

**CONTENT-BASED IMAGE RETRIEVAL OF BLOOD CELLS
BASED ON COLOUR**

ALAGU RATHINA BAIRAVI

**FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA**

2009

**CONTENT-BASED IMAGE RETRIEVAL OF BLOOD CELLS
BASED ON COLOUR**

ALAGU RATHINA BAIRAVI

**DISSERTATION SUBMITTED IN FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER OF
COMPUTER SCIENCE**

**FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA**

2009

Abstract

The research in Content-based image retrieval is developing rapidly. It benefits many other fields, in particular the medical field as the need of having a better way of managing and retrieving digital images has increased.

The aim of this thesis is to create an image retrieval system that will assist hematologist to detect and classify unknown blood cell images speedily and accurately. The algorithm in this study is simple and fast. The problem that arises in this study was difficulty in differentiating one image from the other, as the level of homogeneity of the blood cells is quite high.

The method used in this study was segmentation of blood cell images that was performed based on threshold value automatically selected from a colour histogram. The segmented image is then subtracted with the blood cell images, which are already manually segmented and stored in the database. Images, that has less than 10% difference between automated threshold and manually selected threshold value are retrieved and displayed. Among the 100 images tested, only four images could not be segmented due to the appearance of the images. The number of images that has a difference of less than 10% between manually selected threshold value and automated threshold value are 91 images.

This study of image retrieval of blood cells based on colour generated good results for further prospective study. Additions of more features such as shape, texture, and spatial location will be required to further enhance accuracy of image retrieval of blood cells in the future.

Acknowledgement

This thesis was prepared and completed against great odds and challenges. My first thanks goes to GOD.

There are a number of people that I'm very indebted to for helping me complete this work. Words alone cannot express the thanks I owe to my supervisor, Dr.Woo Chaw Seng, for his help, guidance and support.

I wish to thank my previous supervisor, Ms.Mangalam Sankupellay, and not to forget my second supervisor the late Dr. Prof.Ir.N.Selvanathan, for his help and encouragement.

I would also like to thank my husband, Dr.R.Punithanathan, for his encouragement and helping me to get through it all. Lastly I would like to thank my parents and brother for their advice, support and belief in me.

I should not forget to thank both my sons, Visva and Skandha, for making this as challenging as possible.

TABLE OF CONTENTS

Abstract.....	ii
Acknowledgement	iii
List of Figures.....	vi
List of Tables	viii
CHAPTER I.....	1
INTRODUCTION.....	1
1.1 <i>Background</i>	<i>1</i>
1.2 <i>Motivation and Contribution</i>	<i>4</i>
1.2.1 Statement of Problem.....	5
1.2.2 Purpose and Objectives of Study	6
1.2.3 Scope	7
1.3 <i>Thesis Organization</i>	<i>8</i>
CHAPTER II.....	10
LITERATURE REVIEW.....	10
2.1 <i>Content-based Image Retrieval.....</i>	<i>10</i>
2.2 <i>Visual Features</i>	<i>11</i>
2.2.1 Colour.....	11
2.2.1.1 RGB colour model	11
2.2.1.2 Method of Representation: Colour Histogram.....	11
2.2.2 Texture	12
2.2.3 Shape	13
2.3 <i>CBIR in the Medical Domain.....</i>	<i>14</i>
2.3.1 CBIR Medical Systems	14
2.3.1.1 Image Retrieval in Medical Applications (IRMA)	15
2.3.1.2 I-Browse.....	16
2.3.1.3 The Knowledge-Based Multimedia Medical Distributed Database System	17
(KMeD).....	17
2.3.1.4 PathMaster	18
2.3.1.5 ASSERT.....	18
2.3.1.6 medGIFT	19
2.3.1.7 COBRA.....	20
2.3.1.8 I ² C	20
2.3.2 Image Retrieval in Hematology Study	21
2.3.2.1 Image segmentation	21
2.3.2.2 Thresholding	22
2.3.2.3 Region detection.....	23
2.3.2.4 Feature Extraction	23

2.3.2.4	Relevance feedback system.....	24
2.4	Summary.....	26
CHAPTER III		27
METHODOLOGY.....		27
3.1	<i>Image Acquisition.....</i>	27
3.1.1	Data Sets.....	27
3.1.1.1	Blood cell structure	28
3.1.1.2	Preparation of blood cell smears	28
3.1.1.3	Classes of normal and abnormal blood cell	29
3.2	<i>Image Retrieval Method.....</i>	32
3.3	<i>The Algorithm</i>	33
3.3.1	Framework Overview.....	34
3.3.2	RGB Image to colour components.....	35
3.3.3	Noise Removal.....	36
3.3.4	Find Suitable Threshold Value	38
3.3.5	Segmentation by thresholding each components	40
3.3.6	Logical AND operation (Feature Extraction)	42
3.3.7	Decisions	44
3.4	Summary.....	45
CHAPTER IV.....		46
ANALYSIS OF EXPERIMENTAL RESULTS		46
4.1	<i>Experimental Settings.....</i>	46
4.2	<i>Interim Result Analysis</i>	46
4.3	<i>Experimental Results.....</i>	54
4.4	<i>Discussions.....</i>	66
4.5	Summary.....	69
CHAPTER V		70
CONCLUSIONS		70
5.1	<i>Objectives of Project.....</i>	70
5.2	<i>Strengths of Our Method.....</i>	71
5.3	<i>Future Enhancements.....</i>	72
5.4	<i>Problems Encountered.....</i>	72
5.5	Summary.....	73
Bibliography		75
Appendix.....		86

List of Figures

Fig. 1.1: Diagram for content-based image retrieval system.....	2
Fig. 2.1: I-Browse Architecture	16
Fig. 2.2: KMeD's query interface	17
Fig. 2.3: Interface of medGIFT	19
Fig. 2.4: Image of an eye.....	22
Fig. 2.5: Thresholding at level 160.....	23
Fig. 3.1: Framework of our method.....	33
Fig. 3.2: RGB image of blood cell.....	35
Fig. 3.3: Green component of image.....	35
Fig. 3.4: Blue component of image.....	36
Fig. 3.5: Green component of image after noise removal.....	37
Fig. 3.6: Blue component of image after noise removal.....	37
Fig. 3.7: Image Histogram	38
Fig. 3.8: Pseudocode for finding threshold value	39
Fig. 3.9: Pseudocode for image threshold.....	40
Fig. 3.10: Segmented green component of blood cell.....	41
Fig. 3.11: Segmented blue component of blood cell.....	42
Fig. 3.12: Combined green and blue segmented components.....	43
Fig. 3.13: Combined green and blue segmented components using OR operator.....	43
Fig. 3.14: Pseudocode of image retrieval.....	45
Fig. 4.1: An image from the class of Erythropoiesis.....	46
Fig. 4.2: Green component of the image.....	47
Fig. 4.3: Green component of image after noise removal.....	47

Fig. 4.4: Blue component of image.....	48
Fig. 4.5: Blue component of image after noise removal.....	48
Fig. 4.6: Segmented green component.....	49
Fig. 4.7: Segmented blue component.....	50
Fig. 4.8: Combination of green and blue segments.....	50
Fig. 4.9: Image from the class of Deficiency Anaemis.....	51
Fig. 4.10: Histogram of green component of image.....	52
Fig. 4.11: An image from the class of Deficiency Anaemis.....	53
Fig. 4.12: Segmented green component.....	53
Fig. 4.13: Graph of manual and automated threshold value.....	61
Fig. 4.14: Graph of manual and green threshold value.....	63
Fig. 4.15: Graph of manual and blue threshold value	64
Fig. 4.16: Image A1.7 from the class of Erythropoiesis.....	64
Fig. 4.17: Image A1.8 from the class of Erythropoiesis.....	65
Fig. 4.18: A2.1- Image from the class of Deficiency Anaemis.....	66
Fig. 4.19: A2.6- Image from the class of Deficiency Anaemis.....	66
Fig. 4.20: B1.29- Image from the class of Myeloid cells.....	67
Fig. 4.21: B1.31- Image from the class of Myeloid cells.....	67

List of Tables

Table 3.1: Classes of images.....	27
Table 3.2: Examples of image classes.....	29
Table 4.1: Differences in threshold values for images from the class Erythropoiesis.....	55
Table 4.2: Differences in threshold values for images from the class Deficiency Anaemis.....	56
Table 4.3: Differences in threshold values for images from the class Hemolytic Anaemis.....	57
Table 4.4: Differences in threshold values for images from the class Haemoglobin Disorder.....	58
Table 4.5: Differences in threshold values for images from the class of Red cell Membrane Disorder.....	59
Table 4.6: Differences in threshold values for images from the class Myeloid Cells.....	60
Table 4.7: Images that have less automated threshold value than manual threshold value.....	62
Table 4.8: Percentage of images that has a positive result.....	68

CHAPTER I

INTRODUCTION

This chapter begins with a brief description of the subject. Problem statement, study objectives and limitations of the study are described thoroughly. The motivation and contribution of this study to the relevant field is also stated. The chapter ends with the organization of this thesis.

1.1 Background

Retrieval of digital images in the past has been based on textual annotation for indexing and searching. Images stored in the database are associated with keywords, which are manually annotated to the images. This method is time consuming especially for database with a large number of images. Recently a new method of image retrieval has been introduced. Images are now retrieved based on the content such as the colour, texture, shape, and spatial information.

Content-based image retrieval (CBIR) involves retrieval of images similar to an example query image in terms of some features extracted from the images. It uses visual content such as colour, shape, texture and spatial layout to represent and index the image [1]. CBIR has been an active and challenging research area due to the growth of digital images. CBIR is used in various fields such as:

- Law enforcement: digital forensic investigation, fingerprint, face, DNA matching
- Medical Science: shapes, textures even colour provides critical information in diagnosis of a disease, tumour, abnormalities and others.

- Military
- Advertisement industry
- Earth and space sciences [2]

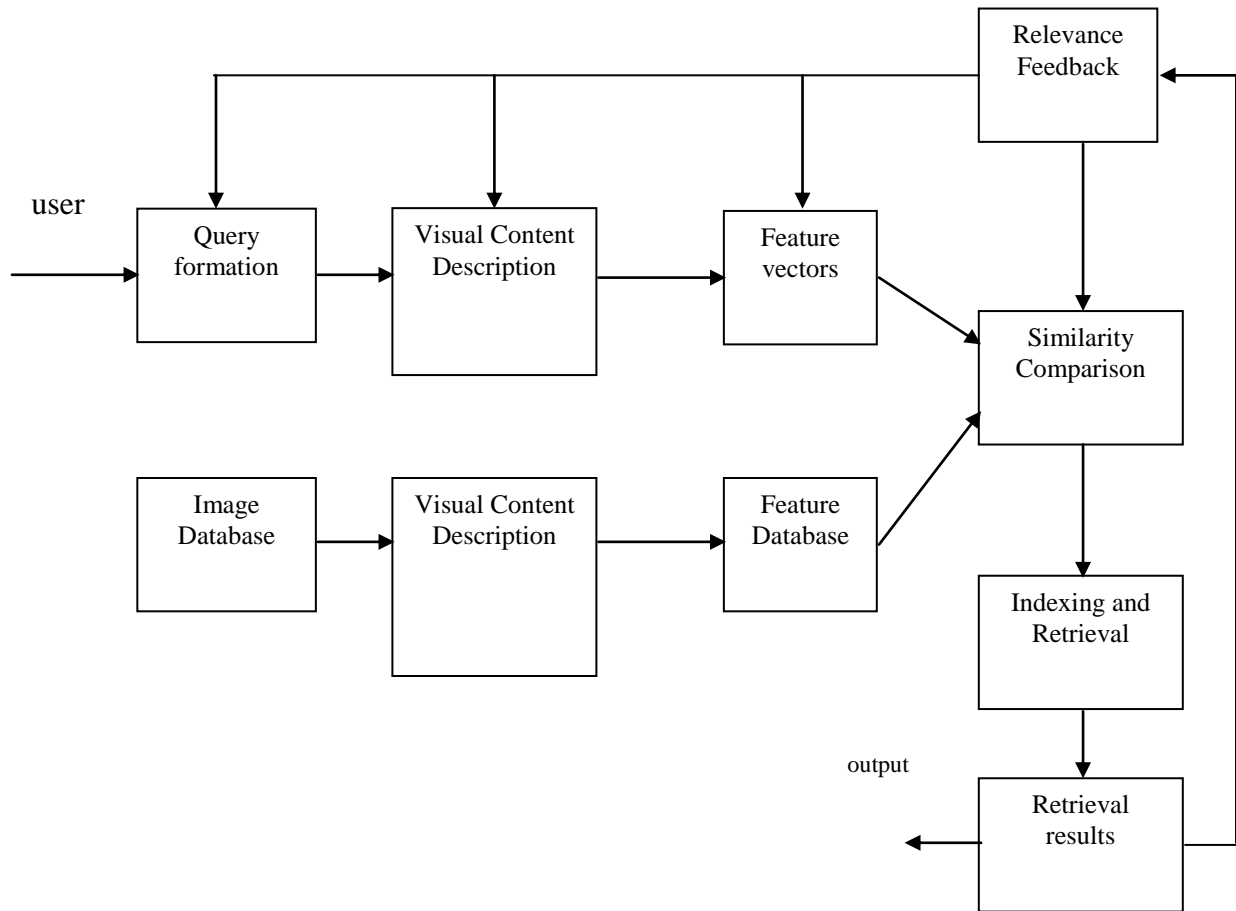


Fig. 1.1: Diagram for content-based image retrieval system [1]

Figure 1.1 depicts the process of a CBIR system. The visual contents of the images are extracted and described by multi-dimensional feature vectors in the CBIR system. These vectors form a database. Users provide a query image into the system and from this query image the CBIR system extracts the feature vectors. The similarities/distances between the feature vector of the query image and the images stored in the database is then calculated. By using an indexing scheme, retrieval is then performed [1]. So each image that is stored in the database has its features extracted and compared to the features of the query image. This involves two major steps:

- **Feature Extraction:** The first step in the process is extracting image features to a distinguishable extent.
- **Matching:** The second step involves matching these features to yield a result that is visually similar.

In the medical field, ever increasing production of digital images for diagnostic purposes has led to more research on image retrieval based on visual contents. With Digital Imaging and Communications in Medicine (DICOM), a standard for image communication has been set and patient information can be stored with the actual image [3]. Several methods from the computer vision and image processing fields already have been proposed for the use in medicine more than ten years ago [4, 5].

The goal of CBIR is different between a medical domain and non-medical domain. The goal in non-medical domain is to retrieve n-images which are closest to the query image while in a medical domain, the goal is to retrieve images which belong to the same disease class as the query image.

CBIR have been implemented in several areas in the medical domain. Some of them are in the areas of dermatology, cytology, pathology, cardiology, histopathology and radiology.

1.2 Motivation and Contribution

The medical institutions can obtain several benefits from implementing automated image retrieval system. It can be used as a foundation to improve service by quick diagnosis and also serve the advantage to reduce cost and increase competitive advantage.

The CBIR system can decrease the time taken to classify blood cell disorders. A hematologist conducts the task of detecting and classifying abnormalities of blood cells. As this procedure is done manually, the time it takes to complete the analysis and to provide a diagnosis is normally longer. With the automated system, the time taken for the classification will be considerably reduced. The inter and intra observer error can also be reduced.

The need for human resources can be reduced with this automated classification of blood cells and hence the reduction in human cost. In centers with high turnover of specimens the reporting time can be reduced.

By reducing the time taken for reporting and also the labour cost, the service provided to the patient is greatly improved. This benefits not only the hematologist but also the whole medical institution.

Teaching efficiency of medical students and postgraduate students can be improved as well through the use of visual access methods. The CBIR system can be used as a learning tool for haematology students for their practical training.

Research in the medical field especially for blood cell classification is increasing. The diagnosis of disease very much depends on the blood cell reporting. Further improvement to the current method is required for a better cell classification and retrieval system.

1.2.1 Statement of Problem

Automated image retrieval of blood cells is not a new venture. There are several research on automated classification and retrieval of different types of blood cell disorders. The intention to overcome the limitations and improve the current techniques of classifying and retrieving cells is always present.

Currently, blood cells are classified into their disease class manually by a hematologist. This is time and energy consuming. So, the image retrieval system will assist hematologist to detect and classify an image of unknown disease class speedily and more accurately.

Problem that is expected in the classification of the blood cell images is the homogeneity of the images. The images are very similar and therefore it is difficult to differentiate one image from the other. This would require more precise descriptors to

distinguish the images. It is hoped that further understanding of the variability of blood cells to the analytical programming will resolve this issues in the future.

1.2.2 Purpose and Objectives of Study

Hematology involves the study of blood cells to detect disorders of the blood. Morphology of cells can provide a wealth of information and assist in making a diagnosis. The purpose of this thesis is to develop a CBIR method for generic blood image classification with low computational cost.

The objectives of this study are:

1. To create a system that is able to classify unknown digital images of blood cell automatically.

The hematologist identifies and classifies blood cells manually. Therefore the system would ease the hematologist's task of classifying unknown blood cells by automating this task.

2. To cater for six types of selected blood images that corresponds to diseases.

The system would be able to classify and retrieve different types of red blood cells. Many literature works have investigated only on a particular blood cell disorder. The systems created are not able to classify and retrieve other types of diseases.

3. To have a simple algorithm that does not require intensive computation.

Most CBIR systems have very complicated methods and this would require intensive computation. For example, in [6], Di Ruberto et. al. investigates on blood cells infected by malarial parasites and the method used is very extensive. This work would benefit more users if it does not require intensive computation. For example, hospitals in third world country with low end machine would benefit from this method.

4. To identify the unknown blood cell images accurately (high precision).

An unknown digital image of blood cell will be identified according to the accurate image class.

5. To retrieve image of blood cell stored in the database that is highly similar to the query image (high recall).

The system will be able to retrieve images stored in the database that is very similar to the query image.

1.2.3 Scope

There are a number of limitations in this study. One of them is the preparation of the digital images. The clarity of the images is highly dependent on the images provided by the haematologist. For the system to be efficient there is a need for high clarity of the images. This will increase the high precision of images.

Another limitation is that the blood cell images are very similar that is the level of homogeneity is very high, therefore they require precise description [7]. There must be an accurate descriptor that is able to distinguish one image from another.

The storage and access methods are important for interactive systems to keep response times at bay. For the system to have less response time, the complexity of the access method has to be given careful consideration. This factor has to be balanced with the descriptor precision.

The result was simulated using 100 images from 6 classes of disease. The method was not tested in the real environment where there would be several images from different classes of disease.

1.3 Thesis Organization

The thesis is divided into five major chapters.

In **Chapter I**, a brief introduction of the thesis is given. Firstly, a short introduction of what is content- based image retrieval is provided. The motivation of the study, problem statement and the scope and limitation of the study is also presented.

Chapter II reviews the literatures relevant to the study. A more elaborate introduction of Content-Based Image Retrieval is given and applications in the different fields are looked into. A number of systems developed and used in the medical field are

also described. Several research have been carried out in this field. The methods applied in some of these research are discussed here.

The method used in this study is discussed in **Chapter III**. Examples of the images used in this study are shown. The step- by- step method used in this study is elaborated.

The experimental results are stated in **Chapter IV**. The settings of the experiment are stated. An analysis of the result is presented with discussion of the success rate of the method used in this study.

Chapter V concludes the thesis. An overview of the study will be discussed. The limitations of the system and further enhancement are stated. Future research and possible applications is also further explored.

CHAPTER II

LITERATURE REVIEW

In this chapter further description of the Content-Based Image Retrieval (CBIR) is provided. The applications of CBIR in the medical domain are also examined with brief introduction of existing medical systems. The chapter ends with a summary of the chapter.

2.1 Content-based Image Retrieval

Image retrieval has been an active area of research. The earlier image retrieval systems were text-based which can be traced back to the 1970s. Images were first annotated with text and text-based database management systems were used to perform image retrieval [8]. The rapid growth in digital images and databases has incurred the need for efficient and accurate retrieval system. Database search based on textual queries can only provide at best a partial solution to the problem. Either the database images are often not annotated with textual descriptions, or the vocabulary needed to describe the user's implied concept does not exist [9]. It was also time consuming to manually annotate each image in the database. Content-based Image Retrieval (CBIR) was introduced in the early 1980s to overcome the disadvantages of text-based image retrieval.

In CBIR images are indexed by their visual contents, such as colour, shape, and texture. There are a few commercial products and experimental prototype systems developed. They are QBIC [10-12], Photobook [13], Virage [14], VisualSEEK [15], Netra [16] and SIMPLIcity [17].

The challenge in CBIR is to retrieve image based on the semantic requirement of user. CBIR can be a failure when there is a semantic gap between user's query image and retrieved image.

2.2 Visual Features

2.2.1 Colour

Colour is the most extensively used visual content for image retrieval. In medical image processing, colour is a very important feature. It is invariant to orientation and scaling [18, 19]. The first step to extract color features is to select an appropriate color space. Several color spaces are available, such as RGB, CMYK, HSV and CIE $L^*u^*v^*$, Munsell, CIE $L^*a^*b^*$ [1, 20]. Image retrieval based on colour features is not very accurate, as images with similar colour do not have similar content. Global colour features computed fails to capture colour distributions or textures within the image.

2.2.1.1 RGB colour model

The Red-Green-Blue (RGB) colour model or RBG colour standard is an additive model in which red, green, and blue (often used in additive light models) are combined in various ways to reproduce other colours. The name of the model and the abbreviation 'RGB' come from the three primary colours, red, green, and blue.

2.2.1.2 Method of Representation: Colour Histogram

Colour histogram is the traditional method of describing colour properties. It provides an efficient representation of colour content. It is also invariant to translation and rotation of

objects[21, 22]. The colour histogram for an image is constructed by counting the number of pixels of each colour [19, 23].

There are two types of colour histograms, Global Colour Histograms (GCHs) and Local Colour Histograms (LCHs). A GCH represents one whole image with a single colour histogram. An LCH divides an image into fixed blocks and takes the colour histogram of each of those blocks [24]. LCHs contain more information about an image but are computationally expensive when comparing images. The GCH is the traditional method for colour based image retrieval. However, it does not include information concerning the colour distribution of the regions of an image. Thus when comparing GCHs one might not always get a proper result in terms of similarity of images.

2.2.2 Texture

Texture of an image is referred as the presence of a spatial pattern that has some properties of homogeneity. There are 3 principal approaches to describe texture. They are:

- Statistical
- Structural
- Spectral

The most common method for capturing texture of an image is Wavelets and Gabor filters. This method is useful for images with homogenous texture.

Fourier power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura feature, Wold decomposition, Markov random field, fractal model are statistical methods and characterize texture by statistical distribution of the image intensity[1].

Texture can be useful to differentiate images, which contain areas of similar colours such as images of sky, leaves or grass.

2.2.3 Shape

Shape features of an image are described after images have been segmented into regions. Methods for shape description are characterized into:

- Boundary-based
- Region based

Boundary-based method includes rectilinear shapes, polygonal approximation, finite element models and Fourier-based shape descriptors. Region based method includes statistical moments. Shape representation of an image should be invariant to translation, rotation and scaling.

2.3 CBIR in the Medical Domain

In the medical field, ever increasing production of digital images for diagnostic purposes has lead to more researches on image retrieval based on visual contents. The objective of using computers to aid diagnosis is to improve the accuracy and increase the consistency of interpretation of medical images.

2.3.1 CBIR Medical Systems

Everyday hospitals and clinics are generating a large number of digital medical images that are useful for diagnosis of diseases, research, and education. These images are from diverse modalities such as, magnetic resonance image (MRI), computerized tomography (CT), ultrasound, x-ray, positron emission tomography (PET), single photon emission computed tomography (SPECT), microscopic pathology and histology images [25]. A brief description of how CBIR is applied in the various medical departments is given in [3]. Currently CBIR systems in medical domain, is limited to a specific organ and modalities and not applicable to other areas [26].

There are several methods proposed on the application of CBIR in different medical domains and various modalities. A methodology for content-based image retrieval of brain tumors is proposed in [27]. The MRI imaging studies of the brain tumor is studied here. Researches on mammogram images is given in [28, 29]. A prototype for content-based image retrieval of pathological images is stated in [30]. The feature types applied here are colour histogram, image texture, Fourier coefficients, and wavelet coefficients. The images used in [31] are cervical and lumbar spine X-rays. Even though there are various studies on

CBIR application in the medical domain, not many have been applied in clinical practice [3].

Some of the existing medical systems and ongoing projects are:

- Image Retrieval in Medical Application (IRMA)
- I-Browse
- The Knowledge-Based Multimedia Medical Distributed Database System (KMeD)
- PathMaster
- Assert
- MedGIFT
- COBRA
- I²C

2.3.1.1 Image Retrieval in Medical Applications (IRMA)

IRMA is targeted on radiological image archive. It aims to automatically classify radiographs based on global features and identify image features [32-34]. The system will retrieve images that are similar to the query image based on a set of features. A coarse classification is first done by texture analysis. Each category is then further sub-divided by segmenting through shape analysis [33].

IRMA is based on a multi-step approach whereby the processes are separated into 7 steps. The steps include categorization using global features, registration in geometry and contrast, feature extraction using local features, feature selection, indexing, identification and retrieval [34].

2.3.1.2 I-Browse

I-Browse is an intelligent content-based image retrieval system in the medical domain. It extracts features combining iconic and semantic content. From these features, it automatically generates textual annotations. The query for this system can either be image or textual query. It also acts as a classifier as the system is able to generate textual annotations of an unknown image [35]. The figure below shows the I-Browse architecture.

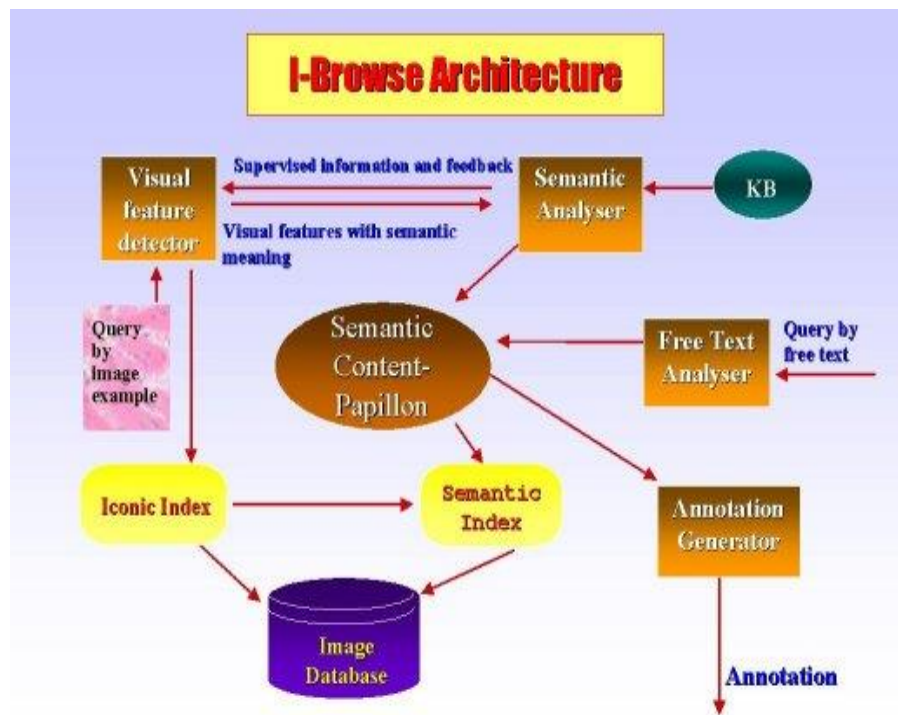


Fig.2.1: I-Browse Architecture [36]

2.3.1.3 The Knowledge-Based Multimedia Medical Distributed Database System (KMeD)

KMeD is a completed research project by University of California Los Angeles (UCLA). The objective of KMeD is to retrieve medical images (e.g. x-ray, MR image) by features such as shape, size, and texture. The content such as spatial relationships among objects is also extracted. The features are stored in a feature and content database [37, 38]. The figure below shows an example of KMeD's query interface.

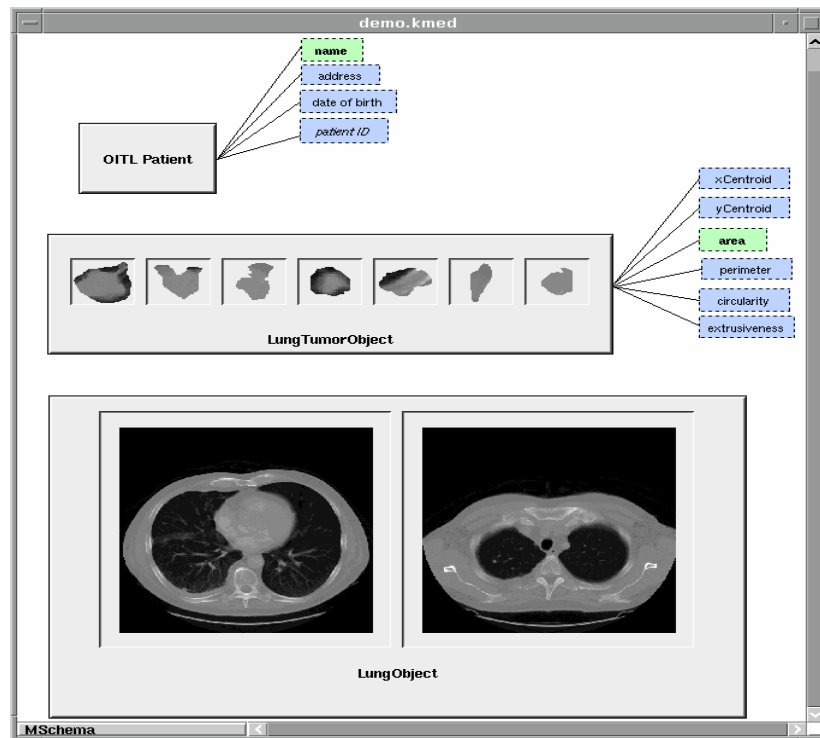


Fig. 2.2: KMed's query interface [38]

2.3.1.4 PathMaster

PathMaster is a program currently under development and refinement for retrieval of cytopathology images. It automatically indexes image of individual cells and their diagnosis.

When user submits a single cell, the features of the cells such as morphology, colour, texture, and spatial relationships are extracted. The features are compared and scored with those stored in the database. A list of cell images sorted from best to worst match is then generated [39].

2.3.1.5 ASSERT

Automated Search and Selection Engine with Retrieval Tools (ASSERT) is a CBIR system targeted to High Resolution Computed Tomographic (HRCT) images of lung with emphysema-type of diseases. It was developed by the School of Electrical and Computer Engineering, Purdue University in collaboration with the Department of Radiology at Indiana University and the School of Medicine at the University of Wisconsin [40].

Radiologists will first highlight pathology bearing regions and then the system will run a query for images with similar regions. This is known as a “physician-in-the-loop” approach [41]. Different image features were used such as co-occurrence statistics, shape descriptors, Fourier transforms and global gray level statistics. It was found that there was an improvement in diagnostic quality when ASSERT was used as a diagnostic aid and was beneficial for less experienced radiologists [42].

2.3.1.6 medGIFT

medGIFT is a image retrieval system that is based on an open source software, GNU Image Finding Tool (GIFT). It extracts local and global features such as textures and grey level descriptors. A variety of images from different modalities (CT, MRI, PET...), photographs and powerpoint slides are used [43]. Images retrieved are arranged to their level of similarity to the query image. A similarity score is displayed under each of the images with the diagnosis. When a user clicks on one of the retrieved images, the corresponding case with textual description and additional images are displayed [44] .

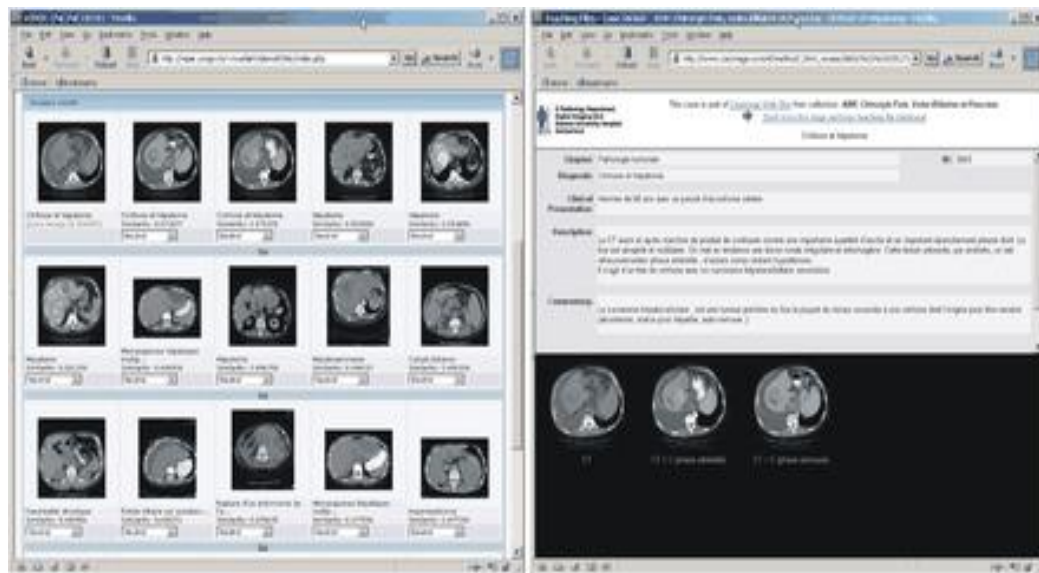


Fig. 2.3: Interface of medGIFT [44]

2.3.1.7 COBRA

Picture Archiving and Communication Systems (PACS) uses textual description to retrieve an image data. Content-Based Retrieval Architecture (COBRA) has further improved the capability of PACS in making diagnosis, research, and training by adding the content feature to image retrieval.

COBRA uses both primitive features such as colour, texture and shape and complex features such as object-based spatial relations and regions of interest [45].

2.3.1.8 I²C

I²C is a system for indexing, storage and retrieval of medical images by content. It was developed based on object-oriented techniques. Segmentation methods and image content description algorithms are encapsulated in image classes. Attributes such as imaging modality, type of study and anatomical characteristics are assigned to image classes. Through this method, the accuracy and efficiency in image retrieval is increased [46].

2.3.2 Image Retrieval in Hematology Study

CBIR provides benefits to hematologist by being able to retrieve similar images to the queried image from the database and therefore improving the quality and efficiency of the diagnosis.

Various studies have been conducted in the area of blood cell classification. Different methods and ideas were applied. Here, four different studies are looked into, to understand the similarity in the methods used and the percentage of success in their method. Basically, there are four main processes identified which are cell segmentation, region detection, feature extraction and, cell classification or retrieval.

2.3.2.1 Image segmentation

In image processing, the aim of segmentation is to locate the objects and boundaries of the image to ease the process of analysing the image. Segmentation involves the process of partitioning the image into multiple regions. Each pixel in the region is similar to some characteristic such as colour, texture or intensity. There are several techniques available for image segmentation. To solve image segmentation problem in the problem domain, these techniques are combined with the domain knowledge, which in this case are blood cells. Cell image segmentation has some problems, which is caused by cell attachments, morphological variation, occlusions, presence of faults and artefacts.

Pan et. al. [21], segments white blood cell from blood and bone marrow cell into cytoplasm and nucleus regions. This is done by image segmentation algorithm and each sub-image represents an individual cell. Matie et. al.[39] segmented Lymphoma cells into

regions of background, cytoplasm, nucleus and nucleolus by creating 2-bit mask overlay of the image. The resulting four levels of gray designate individual regions of interest. While Guangquan et al. [47] applies K-mean clustering method to automatically segment the nucleus regions of leukocytes.

2.3.2.2 Thresholding

Once the images are segmented, a region of interest is determined. In Pan et al., the binary regions corresponding to cytoplasm and nucleus are obtained by histogram thresholding using classical Otsu's method. Thresholding is used to find object within a picture if the brightness level is known. There are two types of thresholding methods, which are uniform and adaptive. In uniform thresholding, pixels above a specified level are set to white while pixels below the specified level is set to black. Optimal thresholding is an advanced level of thresholding whereby a value for threshold is selected that separates object from the background. Otsu's method is a popular method of Optimal thresholding [48]. The figure below shows an image of an eye before thresholding.



Fig. 2.4: Image of an eye [48]

Figure 2.5 shows the image of the eye after uniform thresholding. The value selected for thresholding is 160.



Fig. 2.5: Thresholding at level 160 [48]

2.3.2.3 Region detection

Mattie et al. isolates the region by using binary masks for analysis. Using segmentation mask generates three binary isolation masks, one each for cytoplasm, nucleus, and nucleolus. Guangquan et al. detects single leukocyte region by utilizing thresholding algorithm segmentation and mathematics morphology.

2.3.2.4 Feature Extraction

Feature extraction is a form of dimensionality reduction. Input data is reduced to a set of features, which is called features vector. The transformation of the input data into feature vector is known as feature extraction.

Pan et al, in [21], uses a mixed image representation, which is composed of three normalized histograms. They are:

- RGB colour histogram of individual cell image
- Intensity histogram of cytoplasm
- Intensity histogram of nucleus

The histograms are assembled to a high dimensional vector. The dimensionality is reduced by Kernel Feature Analysis method. In Mattie et al., four types of descriptors are extracted including those of morphology, optical density, texture, and object-relative spatial relationships. The descriptors are calculated separately for each segment. Features extracted in Guangquan et al. are colour, texture, and morphologic features. These features are then normalized.

2.3.2.4 Relevance feedback system

Relevance feedback techniques have been developed to assist users in retrieving the correct images. This technique allows user to choose again from the images retrieved and displayed initially. Users can keep refining the search to the image that is nearest to the query image [49].

Both Pan et al. and Guangquan et al. use relevance feedback system to improve the results of the retrieval system. This system changes the distance metric by re-weighting to features to make relevant images closer. Pan et al. uses an automatic approach to relevance feedback called SVM, which is a type of learning machine.

In [39], the obtained descriptors are used as coordinates to map the characteristics of each cell to a position in feature space. When unknown cell is submitted for evaluation, the distance between its position and the position of other cells in feature space is computed. A

series of cell images and their reports are then generated in an order indicating the distance from the unknown cell.

In [6], Di Ruberto et al. aims to detect parasites infecting red cells in order to evaluate the number of parasites per number of red blood cells. The goal is realized by solving 3 main problems, which are detection of parasites, segmentation of red cells and parasites classification through mathematical operators. The parasites are then detected automatically using colour and size information, which is extracted by morphological approach. A morphological system to segment the blood images is also presented. For parasites classification, two methods, which are used, for feature extraction are morphological approach and colour histogram. In the first approach, classification is solved by analysing shape of parasite while second approach analyses area around the nucleus. Colour histograms describe the colour content of a textured region. In short using morphological analysis has given very good and accurate performance. Di Ruberto et al. report an equally successful result in both methods.

In short, the goal of retrieving cell images most relevant to the query image is realised by executing four identified processes, which are:

- Cell segmentation,
- Region detection,
- Feature extraction
- Cell classification or retrieval.

2.4 Summary

Content Based Image Retrieval (CBIR) is an ongoing research area. It has been applied in different fields. Several CBIR applications are being developed for the medical field to further improve the diagnostic process of diseases. One of the fields in the medical domain is haematology that is the study of blood cells. Classification of blood cells, which is done manually, is automated to ease the job of a hematologist.

The next chapter discusses the methodology used in the study.

CHAPTER III

METHODOLOGY

In this chapter, the methods used for creating a content-based image retrieval system of blood cell images is discussed. The image acquisition algorithm used, pre-processing techniques and samples of some intermediate output are elaborated here.

3.1 Image Acquisition

3.1.1 Data Sets

The data used in this study are images of normal and abnormal blood cells. The images are in Joint Photographic Experts Group (JPEG) format and of varying sizes. For programming purposes, all the images were resized to 280x220 pixels. One hundred images were selected in this study. The images comprised both normal and abnormal red blood cells.

Table 3.1: Classes of images

Class	Number of images
Erythropoiesis	12
Deficiency anaemis	6
Haemolytic anaemis	21
Haemoglobin disorder	18
Red Cell membrane disorder	12
Myeloid cells	31

3.1.1.1 Blood cell structure

Blood is a tissue, which consists of a variety of cells suspended in a fluid medium called plasma. Blood cells can be classified into 3 major functional classes: red blood cells (erythrocytes), white blood cells (leukocytes) and platelets (thrombocytes). Normal red blood cells have a standard size. White blood cells are less than red cells and can be divided into granulocytes and lymphoid cells [50]. The process of blood cell formation is known as haemopoiesis. The haemopoietic system includes bone marrow, liver, spleen, lymph nodes and thymus [51].

3.1.1.2 Preparation of blood cell smears

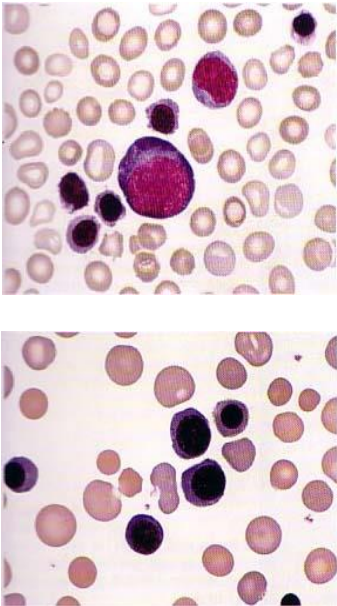
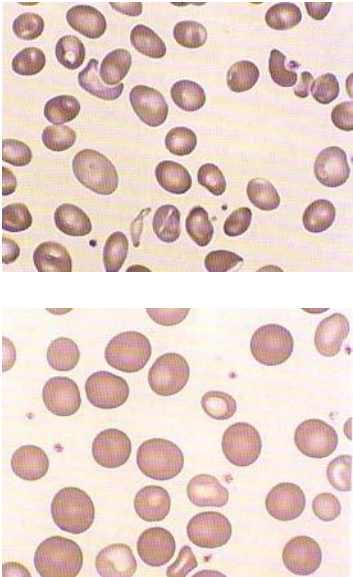
Smearing is needed to obtain images of the blood. An 18 mm by 24 mm clean slides with thickness of 0.7mm is used for smearing purposes. Blood drop (6-7 μ l) is placed at the end of the slide. Another slide is brought into contact with the drop and used to spread evenly. A thin coating is formed on the slide. To be able to observe and recognize the different cells, a process called staining has to be carried out. The slide has to be dried before it can be stained. There are a number of staining method, which is commonly used such as Giemsa, Wright, May-Grünwald and Wright-Giemsa [52].

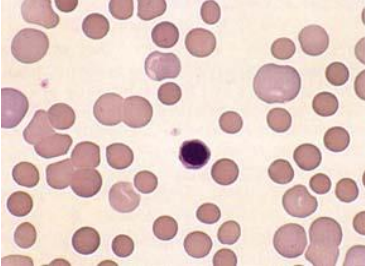
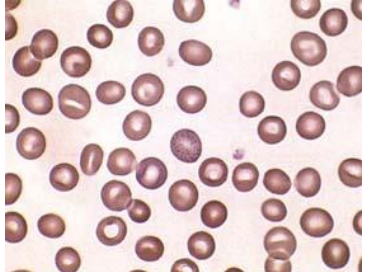
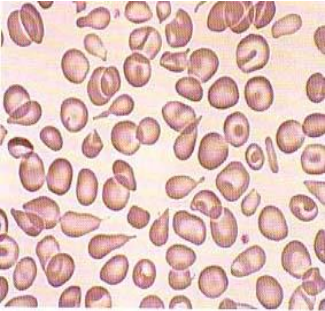
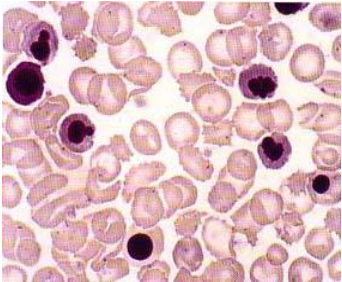
Giemsa stain is an excellent classical blood film stain used on peripheral blood smears and bone marrow. Wright-Giemsa is a dual-use stain for peripheral blood and bone marrow smears. When compared to Giemsa stain, it has a darker result. May-Grünwald is a haematology stain that requires longer staining time but produces a dense coloration. It is somewhat darker than Giemsa and Wright-Giemsa stains [53].

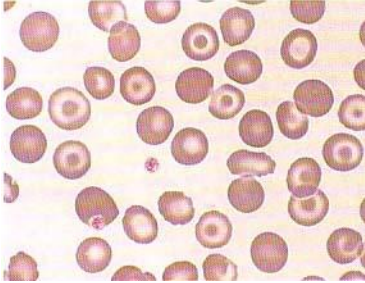
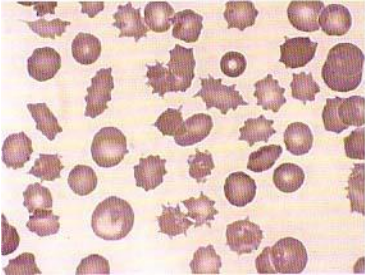
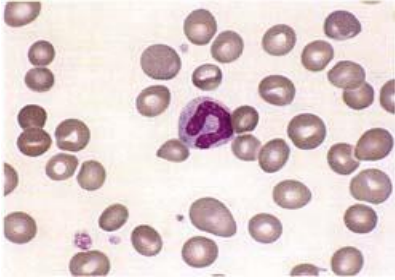
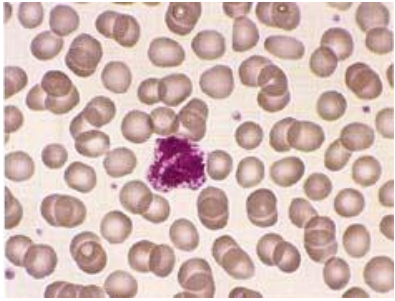
3.1.1.3 Classes of normal and abnormal blood cell

The samples tabled below show examples of various image classes. Short remarks are inserted into each class.

Table 3.2: Examples of image classes

Classes of image	Example of images	Remarks
Erythropoiesis		Process whereby red blood cells are produced in stages in the bone marrow.
Deficiency Anaemias		Anaemia is caused by abnormally low amount of red blood cells. Deficiency includes iron, vitamin B12 and folic acid deficiency.

<p>Haemolytic anaemis</p>	 	<p>A form of anaemia</p> <p>Red blood cells are destroyed and removed before their usual lifespan</p> <p>Bone marrow could not compensate for the loss of erythrocyte which leads to haemolytic anaemis [54]</p>
<p>Haemoglobin disorder</p>	 	<p>It is an inherited blood disorder</p> <p>It affects how oxygen is carried in body</p> <p>There are two types of disorder : Sickle cell and Thalassaemia [55].</p> <p>Hemoglobin disorder causes the shape of a red blood cell to change to a sickle shape</p>

<p>Red cell membrane disorder</p>	 	<p>Membrane of red blood cells are disfigured.</p>
<p>Myeloid cells</p>	 	<p>Cells that is created in bone marrow is known as myeloid cells.</p> <p>Myeloid cells include erythrocytes, granulocytes and platelets [51].</p>

3.2 Image Retrieval Method

Content-Based Image Retrieval (CBIR) systems operate under the query-by-example (QBE) paradigm. An example of the image is presented to the system and the user queries for images that are similar to the given example. Performance of these systems is highly dependent on the properties of the example image.

Firstly, a user uploads the query image through a user interface. The query image is first processed into a binary image, which is then compared with existing images in the image database. A few images that are very similar to the query image is retrieved and displayed.

3.3 The Algorithm

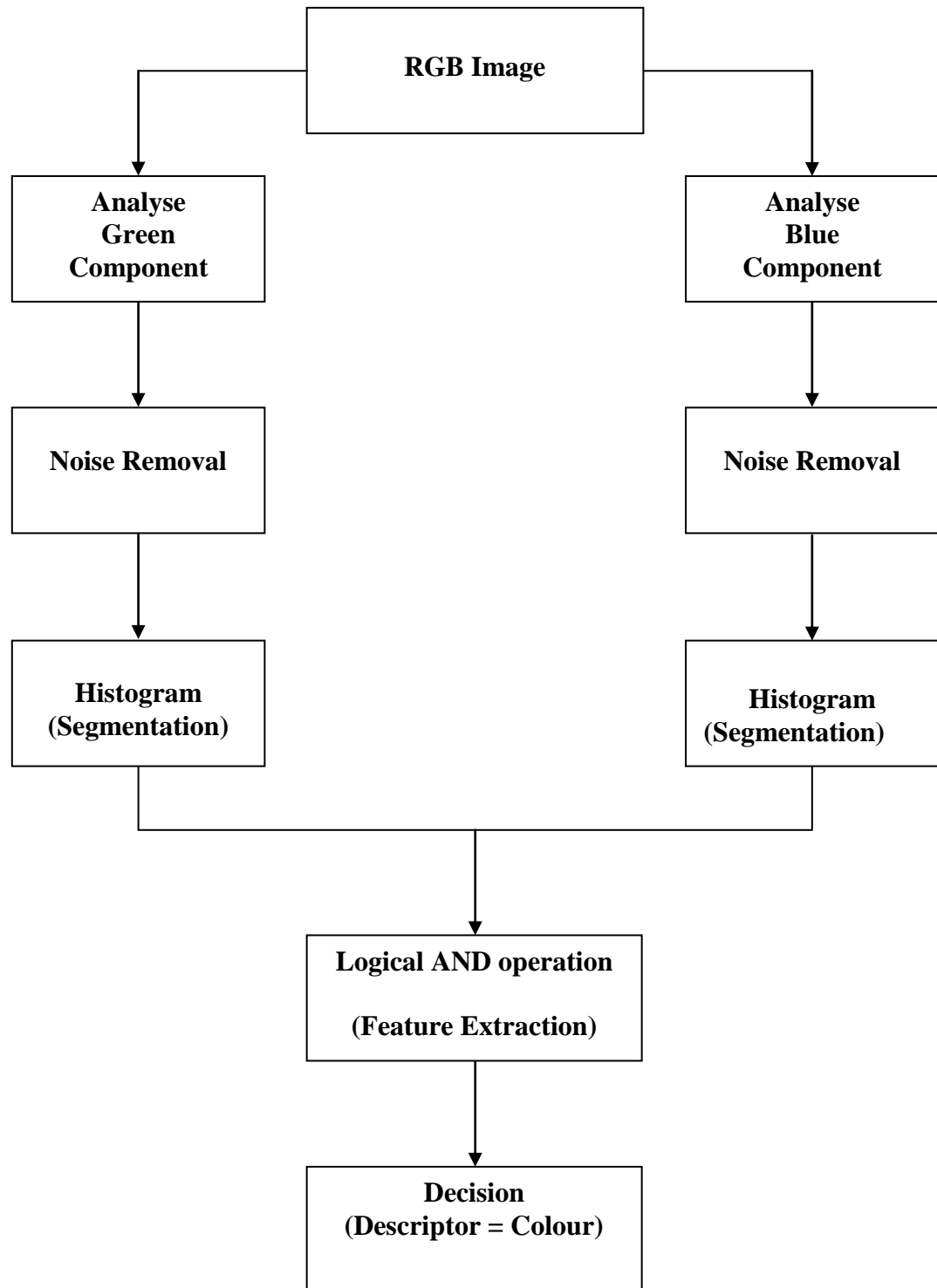


Fig. 3.1: Framework of our method

Above is the framework of the method used in our study. There are 5 major steps:

1. RGB Image to colour components
2. Noise removal
3. Segmentation
4. Logical AND operation (Feature Extraction)
5. Decision

3.3.1 Framework Overview

The RGB image is separated into two different components, which are green and blue. The green and blue components are selected, as red component is too common among all 6 types of blood images. Also from experiment, green and blue components are suitable to differentiate the images. Each of the components goes through the noise removal process to enhance the image clarity. The next step is segmentation of each of the component. Segmentation is realized through thresholding the image. Lastly the segmented components are combined together through the logical AND operation.

3.3.2 RGB Image to colour components

The figure below shows an RGB image of blood cell from the class of Erythropoiesis.

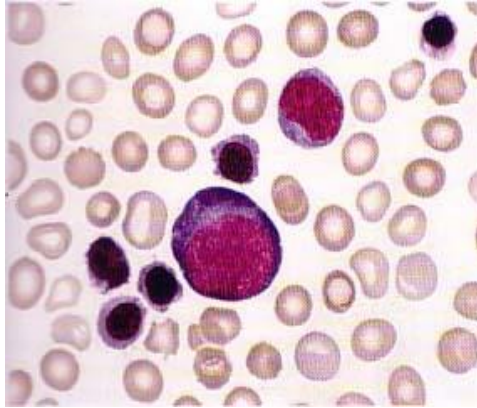


Fig. 3.2: RGB image of blood cell

The blood cell image is then separated into green component and blue component. Figures 3.3 and 3.4 shows the separated green and blue component of the image. These interim results will be further analysed in the next chapter.

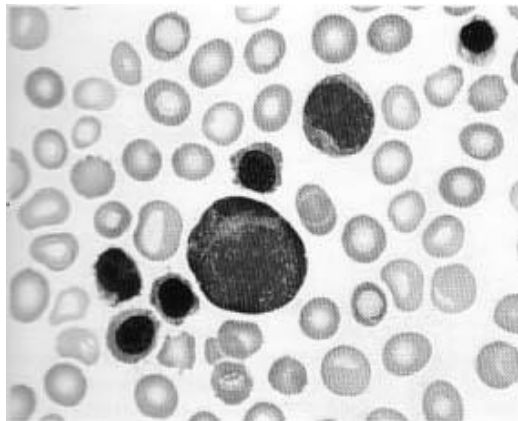


Fig. 3.3: Green component of image

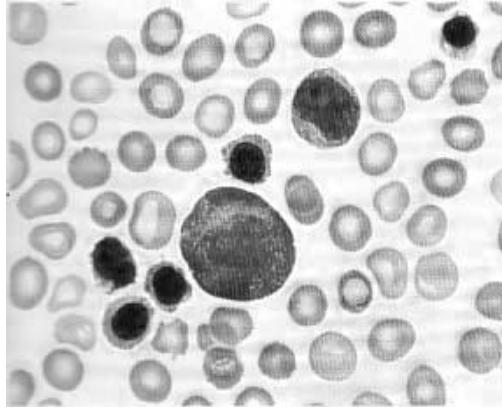


Fig. 3.4: Blue component of image

3.3.3 Noise Removal

Noise is removed from the images to obtain clearer images. This is necessary as digital images acquired may be contaminated by a variety of noise sources. Noise can be removed through linear filtering, median filtering or adaptive filtering. Of these three, the median filter method yields a clearer image.

In median filter, the value of output pixel is determined by the median of the neighbouring pixels [56]. Median filter is better at removing noise without reducing the sharpness of the image. Image is first separated into the green and blue component first. Median filter is then applied to each of the component to remove the noise.

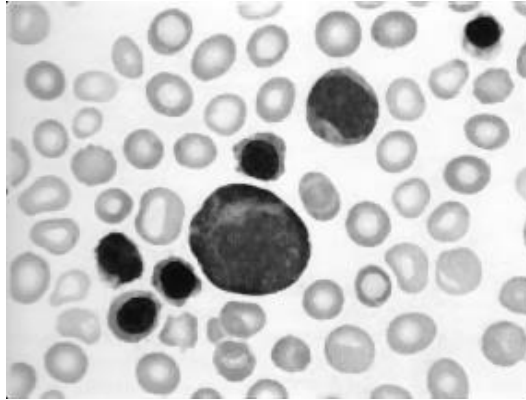


Fig. 3.5: Green component of image after noise removal

Figure 3.5 shows green component of the blood cell image after noise removal. It can be seen that the image is very similar to the original image but appearing to be smooth. The finer details of the image are not lost and appear to be clear.

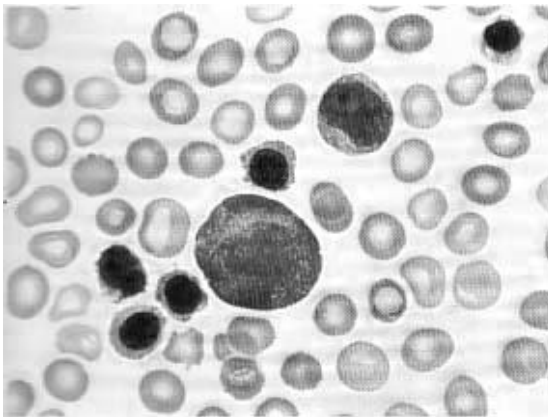


Fig. 3.6: Blue component of image after noise removal

The blue image component is also similar to the original but some details are not present. This can be seen in Fig. 3.6. Once the noise is removed, a threshold value is selected for the images to be segmented.

3.3.4 Find Suitable Threshold Value

A threshold value has to be selected to perform image segmentation. A brightness threshold value is selected and applied to the image. There are several methods to deriving the threshold value. In this study, a histogram is plotted based on the image and threshold value is automatically derived from the histogram. Figure 3.7 below shows an image histogram.

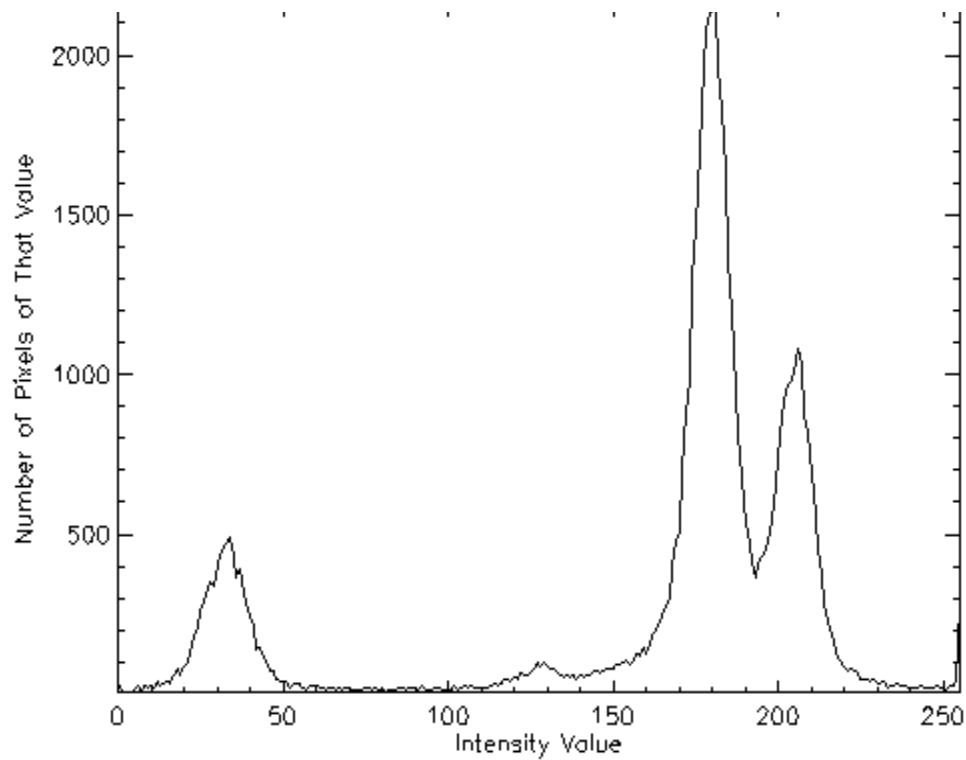


Fig. 3.7: Image Histogram

The threshold value is selected by finding the deepest valley in the histogram. First the largest peak in the histogram is selected. Next the highest peak to the left of the largest peak is then selected. The deepest valley between these two peaks is then chosen. The highest peak to the right of the largest peak is selected. Lastly, the deepest valley between the largest peak and the highest peak to the right is chosen. If the highest peak to the right

of largest peak is greater than highest peak to the left then the threshold value is set to the intensity value of the deepest valley to the right else it is set to the intensity value of the deepest valley to the left. Figure 3.8 shows the pseudocode for finding the threshold value from an image histogram.

```
plot histogram  
  
find largest peak  
  
find highest peak to left of largest peak  
  
find deepest valley between peaks  
  
find highest peak to right of largest peak  
  
find deepest valley between peaks  
  
if highest peak to right < highest peak to left then  
    set threshold value to intensity value of deepest  
    valley to right  
  
else  
    set threshold value to intensity value deepest  
    valley to left  
  
endif
```

Fig. 3.8: Pseudocode for finding threshold value

Once noise is removed from the green and blue component of the image, a histogram is plotted for each of the components. From the histograms, the threshold value is selected and applied for both of the components. The output binary image has values of 0 (black) for all pixels in the input image with luminance less than the threshold value and 1 (white) for all the other pixels. A pseudocode for this is given in Fig. 3.9.


```
get image dimension  
set [i,j] as image dimension  
create an empty black image  
select threshold value from image histogram  
if image[i,j] is greater than threshold value then  
    set image[i,j] to 1  
else  
    set image[i,j] to 0  
end if
```

Fig. 3.9: Pseudocode for image threshold

3.3.5 Segmentation by thresholding each components

In image processing, the aim of segmentation is to locate the objects and boundaries of the image to ease the process of analysing the image. Segmentation involves the process of partitioning the image into multiple regions. Most pixels in a region are similar on some characteristics such as colour, texture, intensity or gray level.

There are several techniques available for image segmentation. To solve image segmentation in the problem domain, these techniques are combined with the domain knowledge, which in this case are blood cells. Cell image segmentation has some problems, which is caused by cell attachments, morphological variation, occlusions, presence of faults and artefacts.

There are several techniques to segmentation. In our method, a thresholding value was selected for image segmentation. This value is automatically selected using our algorithm. Details of this step are described in section 3.3.4. Each of these components is then segmented based on the threshold value obtained from plotted histograms.

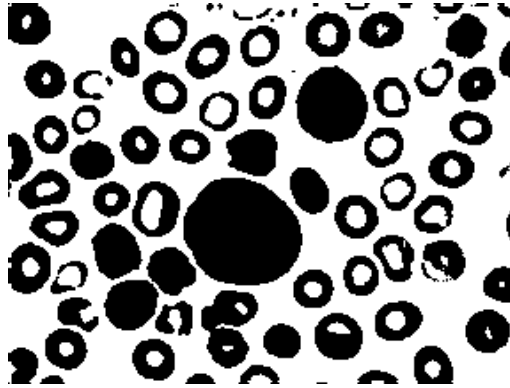


Fig 3.10: Segmented green component of blood cell

Figure 3.10 shows the segmented green component of the blood cell image. From the figure, it can be seen that the segmented green component has clear boundary of cells and distinctive objects are easily segmented.

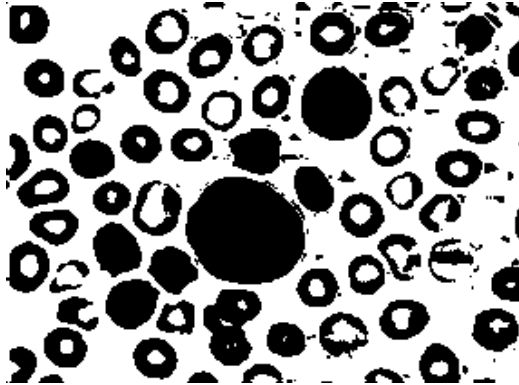


Fig 3.11: Segmented blue component of blood cell

The segmented blue component shows many small dots as depicted in Fig. 3.11. It presents a noisy outcome.

3.3.6 Logical AND operation (Feature Extraction)

Each of the green component and blue component is segmented and the resulting binary image is combined using the AND operator. The operator used is AND instead of OR in order to exclude unnecessary noise and retrieve only the major features. Figure 3.12 below shows the combined green and blue segmented components using the logical AND operator. If the blue and green components are combined using logical OR operation, then the results would have too much noise. An example of such case is presented in Fig. 3.13.

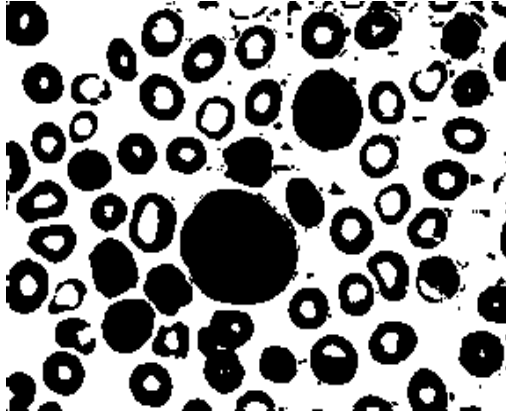


Fig 3.12: Combined green and blue segmented components

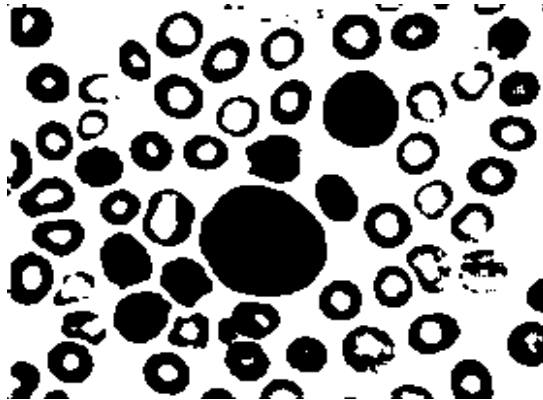


Fig 3.13: Combined green and blue segmented components using OR operation

Feature extraction is a method of capturing the visual content of images for retrieval purposes. The goal of feature extraction is to reduce image dimension while ensuring that the image is still described accurately without sacrificing important information. The feature extracted here is the colour of an image in particular the green and blue intensities of the image.

After examining the histogram of RGB components, we found that green and blue form a major part of the images. The red component of the image was not included because of two reasons:

1. Combining only two components instead of three will reduce computation.
2. The result of combining this two component is equivalent to combining all three components therefore it can be ignored. A visual check on all images revealed that majority of them have “reddish” or orange colour as a major component. This means red component is not helpful in differentiating one image from another image.

3.3.7 Decisions

Query image that is entered is automatically segmented and compared with binary images in a database. A comparison is achieved by subtracting this segmented query image with the binary image in the database. Images are subtracted to find the differences between the query image and the image that should be retrieved. Only results, which differ less than 10 % of the image size, is retrieved and displayed to ensure relevant images are retrieved. This step of subtraction and selecting difference of 10% is to ensure that images retrieved are highly similar when visually compared with the query image. Fig. 3.14 shows the pseudocode for the image retrieval.

```
read image  
  
convert rgb image to binary image  
  
read image query  
  
subtract image query with binary image  
  
get image dimension  
  
if result of subtraction is less than 10% of  
image dimension then  
  
    yes = display retrieved image  
  
    else  
  
        no
```

Fig 3.14: Pseudocode of image retrieval

3.4 Summary

In our CBIR method, an image of the blood cell is first separated into different colour components and each of the components is segmented. The results are binary images that are then compared with the binary images already stored in the database. The query binary image is subtracted from the binary image stored in the database. Only results, which yield less than 10 percent difference of the image size, is retrieved and displayed.

CHAPTER IV

ANALYSIS OF EXPERIMENTAL RESULTS

In this chapter, the experimental results, settings and analysis are presented and discussed.

4.1 Experimental Settings

One hundred images were used for the study with different blood disorders. The images can be classified into 6 classes of blood disorders. Each of the images was resized to same dimensions and they are all in JPEG format. The experiment was conducted on an AMD Sempron™ 2200+ machine with a processor speed of 1.49 GHz, 480MB RAM, hard disk size of 80GB and Windows XP operating system. The software utilized for this study is MATLAB as it has built in library functions for image processing.

4.2 Interim Result Analysis

See figure 3.1, the framework for the method applied, in the previous chapter.

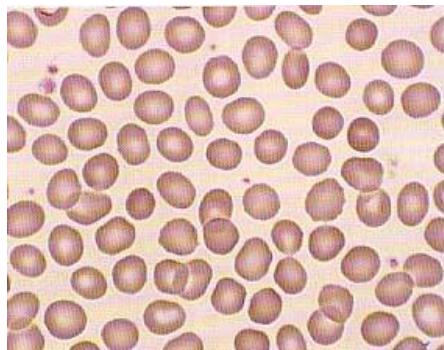


Fig. 4.1: An image from the class of Erythropoiesis

The above figure 4.1 shows a blood cell image before any processing is carried out. The image shows red blood cell of the class Erythropoiesis. There is a clear boundary between most of the cells with some cells overlapping. The colour does not show a deep red but it light brown. The image is then separated into green component and blue component.

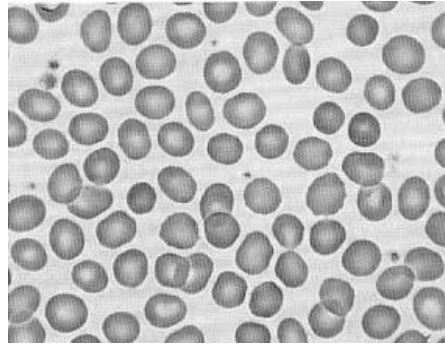


Fig. 4.2: Green component of the image

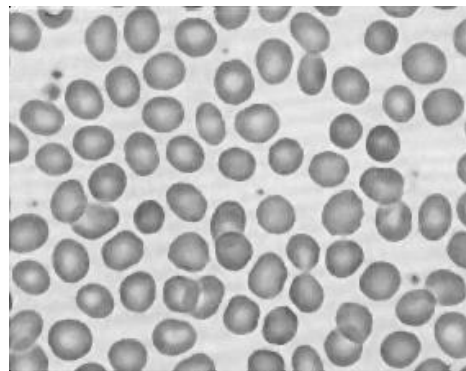


Fig. 4.3: Green component of image after noise removal

Figure 4.2 shows the green component of the image and figure 4.3 shows the green component of the image after noise removal. Noise is information which is not required that contaminates the image. Noise may be present due to the image acquisition process.

So, the elimination of noise is needed to minimise errors. The noise was removed through median filter.

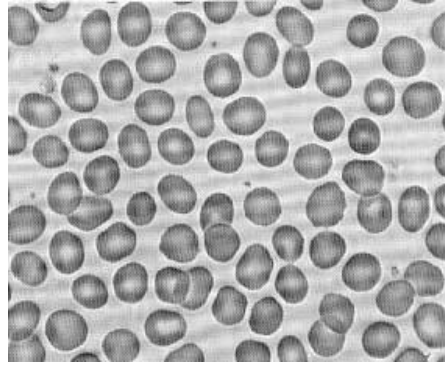


Fig. 4.4: Blue component of image

The above figure shows the blue component of the image. Here, it can be seen that the cells appear darker nearing the border and with some level of noise. This suggests that the cells contain more of blue component than green. The blue component also undergoes noise removal and the result is shown in the figure below.

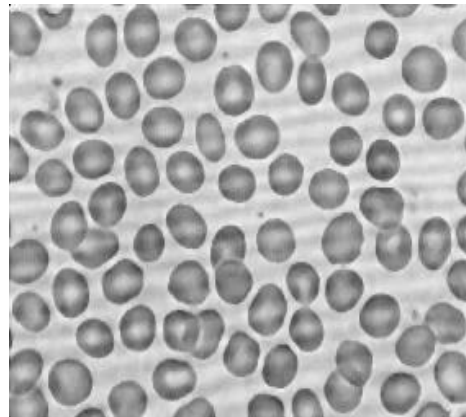


Fig. 4.5 : Blue component of image after noise removal

The noise removal process did not yield any clear result. The appearance of the image does not show any changes except to have a blurred appearance.

The green component is then segmented using a threshold value obtained from a plotted histogram. The figure 4.6 below shows the segmented green component of the image.

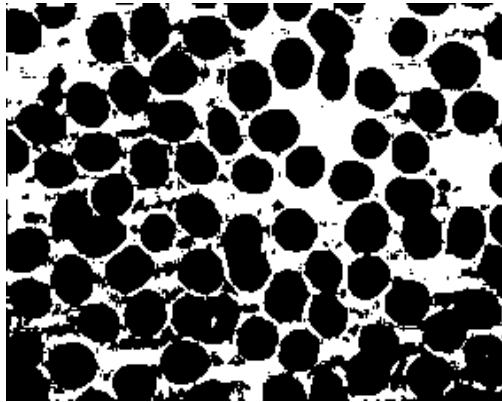


Fig. 4.6 : Segmented green component

The segmented green does not yield a clear result. All the cells have no clear boundaries and with high level of noise.

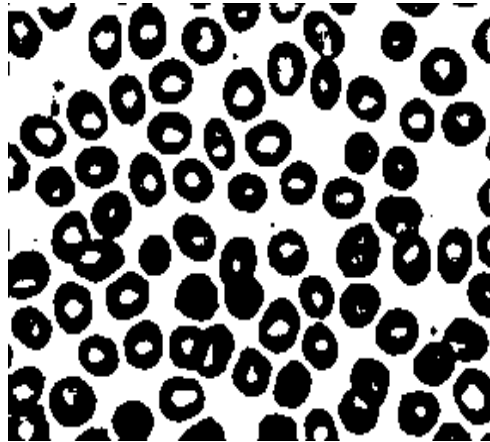


Fig. 4.7 : Segmented blue component

The segmented blue component is clearer with no noise. Cell boundaries are clear. The segmented green and blue components are combined together by using the logical AND operator. The result shows no clear boundary between overlapping cells and too many small (noise) dots. Figure 4.8 shows combination of green and blue components.

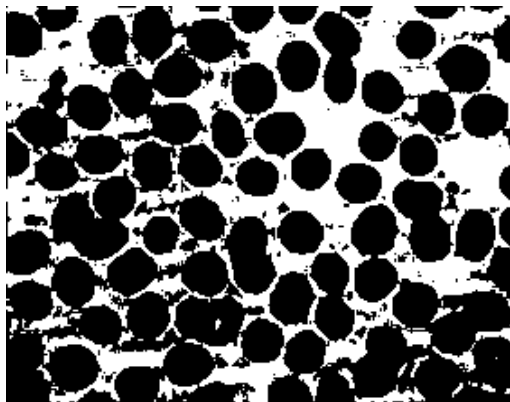


Fig. 4.8 : Combination of green and blue segments

Among the 100 images, 4 images could not be segmented. Two out of the 4 images are discussed here. Both images are from the class of Deficiency Anaemis. The following paragraphs analyse the 2 special cases that could not be segmented correctly.

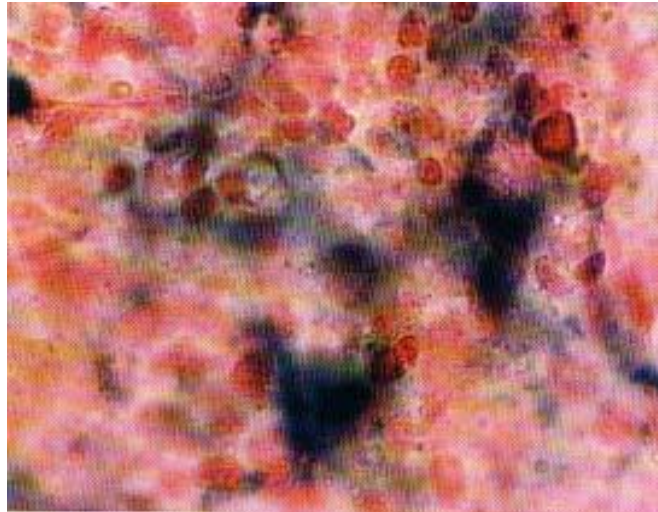


Fig. 4.9: Image from the class of Deficiency Anaemis

The figure 4.9 shows an image from the class of Deficiency anaemis. The image was separated into the green component and blue component and plotted the histogram. The figure below shows the histogram for the green component.

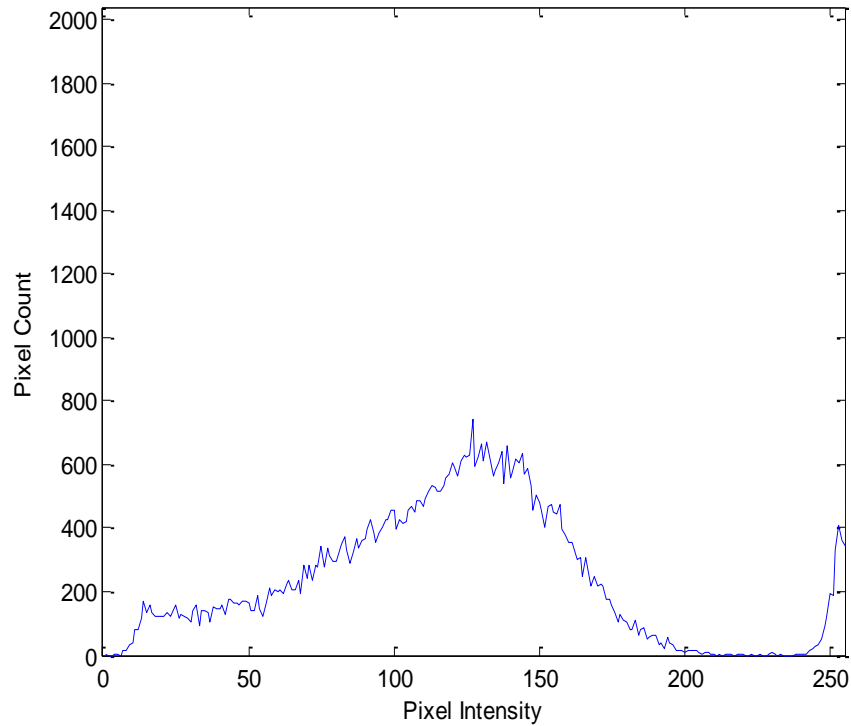


Fig 4.10 : Histogram of green component of image

Segmenting the green component based on the threshold value from the above histogram only gives a white background. The blue component also gives a similar histogram and segmentation yields a white background as well. These results mean the segmentation do not produce any cells.

Another image, which, could not be segmented is also from the class of Deficiency Anaemis.

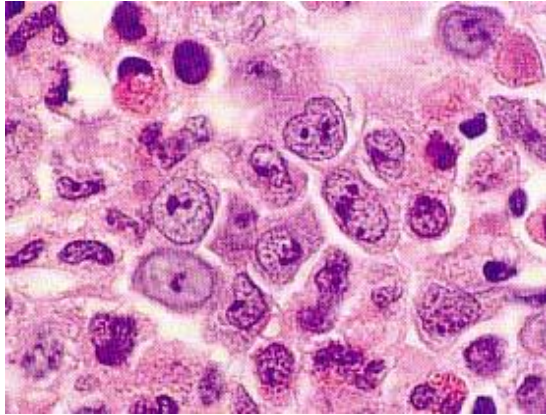


Fig. 4.11 : An image from the class of Deficiency Anaemis

The figure below shows the segmented green component.



Fig. 4.12: Segmented green component

The segmented blue component only shows a black background. These problems arise due to the homogenous content of the 2 images. They do not have distinctive cells for automated segmentation. So, segmentation could not be performed on both these images.

4.3 Experimental Results

The method used in this study was to separate the image into green and blue components. Noise acquired during the image acquisition process was first removed by using a median filter. Each of these image components is then segmented based upon automatically derived threshold value from the image histogram. The resulting image segment of each component is then combined using AND operator.

This segmented image was then compared with images that are manually segmented by human and stored in the database. Images, which, yields less than 10 percent difference of the image size were retrieved and displayed. A ten percent difference was chosen here, as the results will be very similar to the manually segmented results. The automated segmentation results are as good as human processed results.

The threshold value was chosen manually for image segmentation. The process is done by visually determining a suitable threshold value. This was compared with the automatically derived threshold value, which comprises of the threshold value for the green component of the image, blue component of the image and the average of both components of the image.

Given below are the differences of the automatic threshold value and manually chosen threshold value. It also shows the threshold value for the green and blue component and also whether the difference between automated and manual threshold value falls under 10%.

Table 4.1: Differences in threshold values for images from the class Erythropeiosis

Image Number	Manual threshold	Green component threshold	Blue component threshold	(Auto) Average threshold	Difference <10% (Y/N)
A1.1	0.6	0.79	0.8	0.79	Y
A1.2	0.6	0.74	0.77	0.75	Y
A1.3	0.6	0.7	0.65	0.68	Y
A1.4	0.7	0.84	0.83	0.84	Y
A1.5	0.7	0.84	0.82	0.83	Y
A1.6	0.6	0.76	0.69	0.73	Y
A1.7	0.4	0.22	0.7	0.46	Y
A1.8	0.4	0.22	0.52	0.37	N
A1.9	0.3	0.54	0.7	0.63	N
A1.10	0.4	0.75	0.86	0.8	N
A1.11	0.5	0.27	0.6	0.44	N
A1.12	0.4	0.27	0.6	0.44	Y

In the above table, the threshold values are mostly between 0.63 and 0.84. Images A1.7 and A1.8 have lesser green components and higher blue components. The automated and manual threshold value for image A1.9 shows a very high difference. The other images show a difference of between 10% to 20% in the automated and manual threshold value. Out of 12 images, 8 images shows a difference of less than 10% between the automated and manual threshold values that is 66.7%.

Table 4.2: Differences in threshold values for images from the class Deficiency Anaemis

Image Number	Manual threshold	Green component threshold	Blue component threshold	Average threshold	Difference <10% (Y/N)
A2.2	0.7	0.78	0.69	0.73	Y
A2.3	0.8	0.82	0.78	0.8	Y
A2.4	0.7	0.79	0.75	0.77	Y
A2.5	0.6	0.59	0.68	0.64	Y

There is a stable value from 0.64 to 0.80 for automatic threshold. Two images, A2.1 and A2.6 are not included in the table above, as it did not give any threshold value. All 4 images' has a less than 10% difference between automated and manual threshold value.

Table 4.3: Differences in threshold values for images from the class Hemolytic Anaemis

Image Number	Manual threshold	Green component threshold	Blue component threshold	Average threshold	Difference <10% (Y/N)
A3.1	0.7	0.79	0.76	0.78	Y
A3.2	0.8	0.87	0.84	0.85	Y
A3.3	0.6	0.77	0.75	0.76	N
A3.4	0.7	0.8	0.79	0.79	Y
A3.5	0.4	0.3	0.62	0.46	Y
A3.6	0.7	0.79	0.78	0.78	Y
A3.7	0.7	0.72	0.69	0.71	Y
A3.8	0.7	0.79	0.79	0.79	Y
A3.9	0.7	0.78	0.74	0.76	Y
A3.10	0.7	0.78	0.74	0.76	Y
A3.11	0.7	0.8	0.78	0.79	Y
A3.12	0.8	0.84	0.8	0.82	Y
A3.13	0.7	0.8	0.77	0.79	Y
A3.14	0.7	0.8	0.77	0.79	Y
A3.15	0.7	0.78	0.75	0.76	Y
A3.16	0.8	0.82	0.81	0.82	Y
A3.17	0.7	0.79	0.8	0.79	Y
A3.18	0.8	0.85	0.78	0.82	Y
A3.19	0.7	0.8	0.78	0.79	Y
A3.20	0.8	0.82	0.8	0.82	Y
A3.21	0.8	0.89	0.82	0.85	Y

Based on the above table from the class of Hemolytic Anaemis, the manual threshold values for all 21 images are less than the automated values. The difference between manual and automated is less than 10% in 20 images.

Table 4.4: Differences in threshold values for images from the class Haemoglobin disorder

Image Number	Manual threshold	Green component threshold	Blue component threshold	Average threshold	Difference <10% (Y/N)
A4.1	0.7	0.82	0.78	0.79	Y
A4.2	0.7	0.79	0.78	0.79	Y
A4.3	0.8	0.84	0.85	0.85	Y
A4.4	0.8	0.84	0.85	0.84	Y
A4.5	0.7	0.79	0.78	0.79	Y
A4.6	0.7	0.78	0.72	0.75	Y
A4.7	0.7	0.8	0.75	0.78	Y
A4.8	0.6	0.67	0.69	0.67	Y
A4.9	0.7	0.77	0.72	0.74	Y
A4.10	0.7	0.72	0.67	0.69	N
A4.11	0.8	0.86	0.83	0.84	Y
A4.12	0.7	0.77	0.72	0.74	Y
A4.13	0.7	0.76	0.74	0.75	Y
A4.14	0.7	0.74	0.68	0.71	Y
A4.15	0.7	0.72	0.73	0.72	Y
A4.16	0.7	0.79	0.76	0.77	Y
A4.17	0.7	0.77	0.68	0.73	Y
A4.18	0.7	0.74	0.7	0.72	Y

Based on the table above, manual threshold value of the images is lesser than automated threshold value but does not show a marked difference. Only one image, which is A4.10 shows difference compared to other images. Out of 18 images, 17 images yield a difference less than 10% between manual and automated threshold.

Table 4.5: Differences in threshold values for images from the class of Red cell Membrane Disorder

Image Number	Manual threshold	Green component threshold	Blue component threshold	Average threshold	Difference <10% (Y/N)
A5.1	0.7	0.76	0.71	0.73	Y
A5.2	0.8	0.79	0.84	0.81	Y
A5.3	0.7	0.82	0.81	0.82	N
A5.4	0.8	0.83	0.81	0.82	Y
A5.5	0.7	0.8	0.74	0.77	Y
A5.6	0.7	0.77	0.71	0.74	Y
A5.7	0.7	0.79	0.75	0.77	Y
A5.8	0.7	0.79	0.73	0.76	Y
A5.9	0.7	0.78	0.76	0.77	Y
A5.10	0.8	0.84	0.83	0.84	Y
A5.11	0.8	0.84	0.83	0.84	Y
A5.12	0.7	0.79	0.75	0.77	Y

For the images from the class of red cell membrane disorder, the threshold values for both manual and automated does not show a huge difference. Only image A5.3 shows more difference between manual and automated threshold values in comparison with other images. Among 12 images, there are 11 images that have differences of less than 10% between manual and automated threshold.

For images from the class of Myeloid cells, 3 images out of 31 images have manual threshold value more than the automated value. In the image B1.27, the green component shows a lesser value than the blue component by 51%. Only two images have the same value for automated and manual threshold that is B1.3 and B1.10. Out of the 31 images, 22 images have manual and automated threshold values that are less than 10%, which is 71%. Two images, B1.29 and B1.31 did not obtain any values for automated thresholding and this is shown as dash sign in the relevant columns.

Table 4.6: Differences in threshold values for images from the class Myeloid cells

Image Number	Manual threshold	Green component threshold	Blue component threshold	Average threshold	Difference < 10% (Y/N)
B1.1	0.7	0.75	0.72	0.74	Y
B1.2	0.7	0.76	0.75	0.76	Y
B1.3	0.8	0.79	0.82	0.8	Y
B1.4	0.7	0.82	0.77	0.79	Y
B1.5	0.7	0.79	0.79	0.79	Y
B1.6	0.7	0.79	0.79	0.79	Y
B1.7	0.8	0.84	0.83	0.83	Y
B1.8	0.8	0.82	0.8	0.81	Y
B1.9	0.7	0.79	0.76	0.78	Y
B1.10	0.9	0.9	0.9	0.9	Y
B1.11	0.7	0.34	0.86	0.6	N
B1.12	0.5	0.34	0.92	0.63	N
B1.13	0.8	0.84	0.82	0.83	Y
B1.14	0.7	0.76	0.72	0.74	Y
B1.15	0.7	0.77	0.76	0.77	Y
B1.16	0.7	0.81	0.79	0.8	N
B1.17	0.7	0.74	0.71	0.72	Y
B1.18	0.7	0.82	0.77	0.79	Y
B1.19	0.8	0.86	0.8	0.83	Y
B1.20	0.8	0.84	0.86	0.85	Y
B1.21	0.7	0.82	0.8	0.81	N
B1.22	0.7	0.79	0.76	0.78	Y
B1.23	0.7	0.83	0.76	0.79	Y
B1.24	0.7	0.78	0.78	0.78	Y
B1.25	0.8	0.84	0.84	0.84	Y
B1.26	0.4	0.37	0.58	0.48	Y
B1.27	0.6	0.27	0.78	0.53	N
B1.28	0.5	0.3	0.51	0.41	N
B1.29	0.6	-	-	-	-
B1.30	0.7	0.84	0.8	0.82	N
B1.31	0.7	-	-	-	-

A line graph is plotted based on the manual and automated threshold values to show the difference between both the values. Four images did not produce any threshold values. For the purpose of plotting the graph a test value was entered for those images. Figure 4.13

below shows the graph for manual, automated, green, and blue threshold values. From the graph below, it can be seen that the manually chosen threshold value is lesser than the automated threshold value.

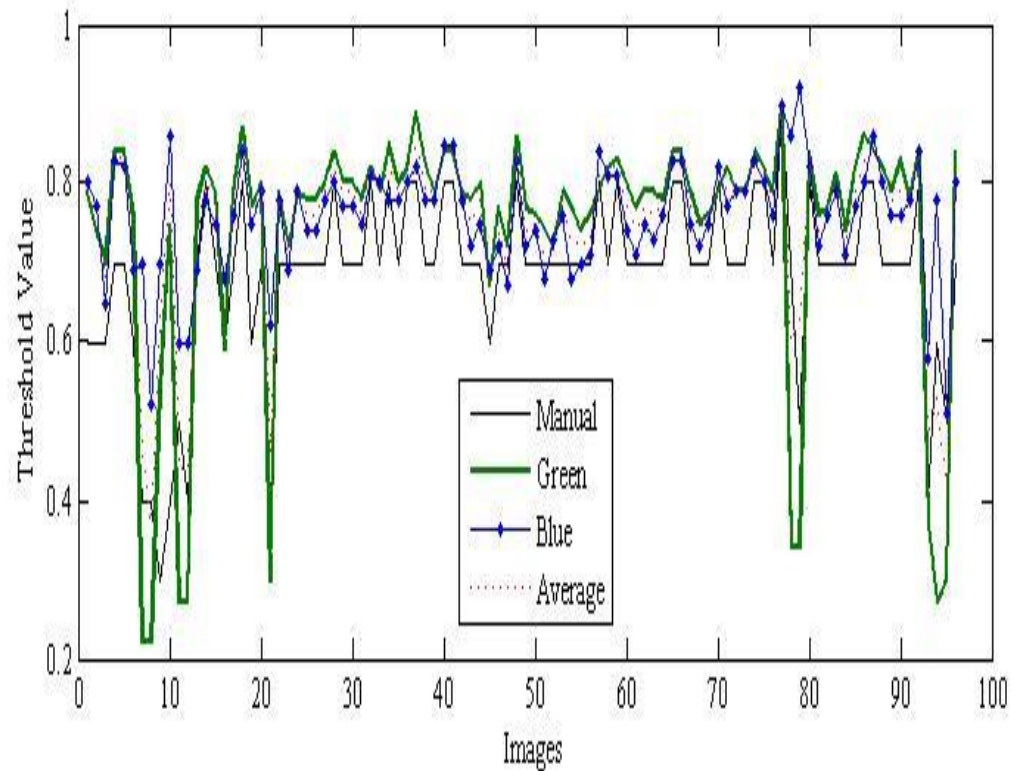


Fig 4.13: Graph of manual and automated threshold value

Different classes of image too yield different results. Images A1.1 to A1.12 which falls under the class of Erythroipoiesis, has a lesser threshold value than the automated threshold value except for one image which is A1.8.

The table below shows the blood cell images from the different classes that have a manual threshold value more than automated threshold value.

Table 4.7: Images that have less automated threshold value than manual threshold value

Image Number	Image Class	Manual threshold	Automated threshold
A1.8	Erythropoiesis	0.4	0.37
A4.10	Haemoglobin Disorder	0.7	0.69
B1.11	Myleoid cells	0.7	0.6
B1.27	Myleoid cells	0.6	0.53
B1.28	Myleoid cells	0.5	0.41

The green component and blue component's threshold value too have to be compared with the manually chosen threshold value. Graphs for comparing the green and blue component's threshold value with the manual threshold value is also plotted separately and shown in figure 4.14 and 4.15 respectively.

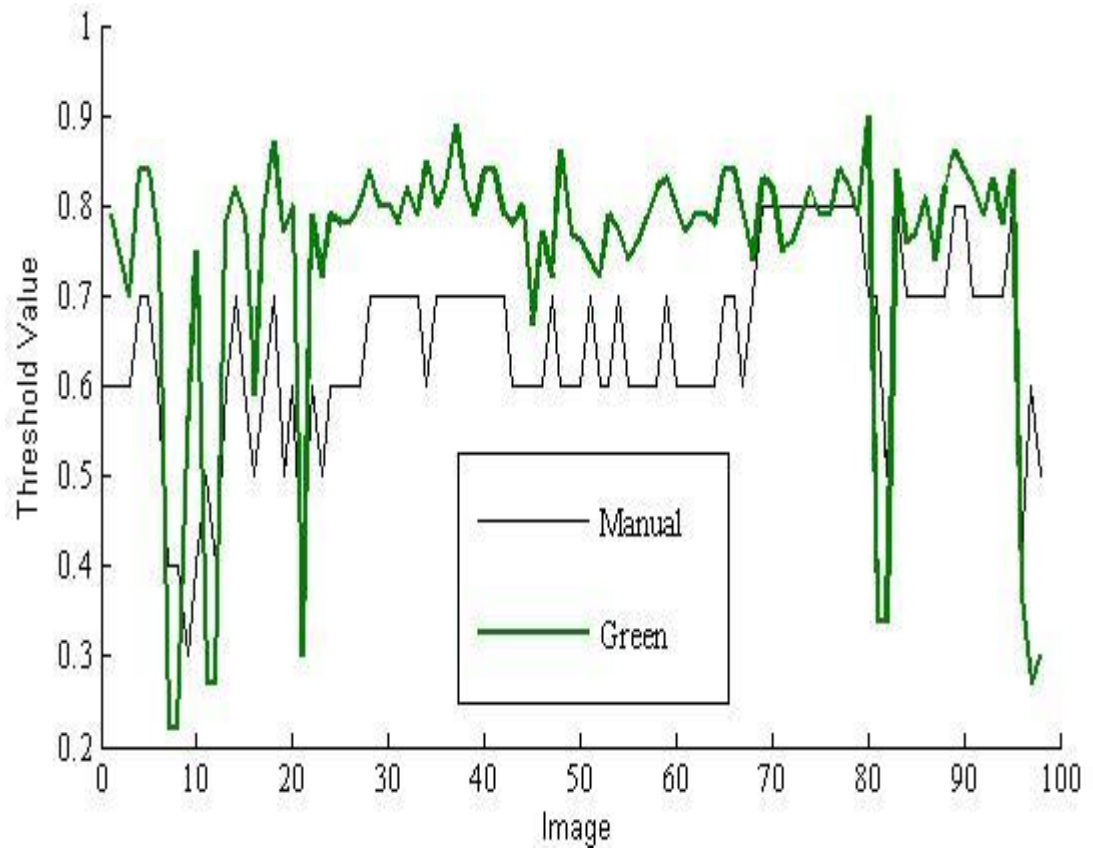


Fig 4.14: Graph of manual and green threshold value

Overall the threshold value for the blue component is higher than the manual threshold value when compared to the value for green component, which is less than the manual threshold value.

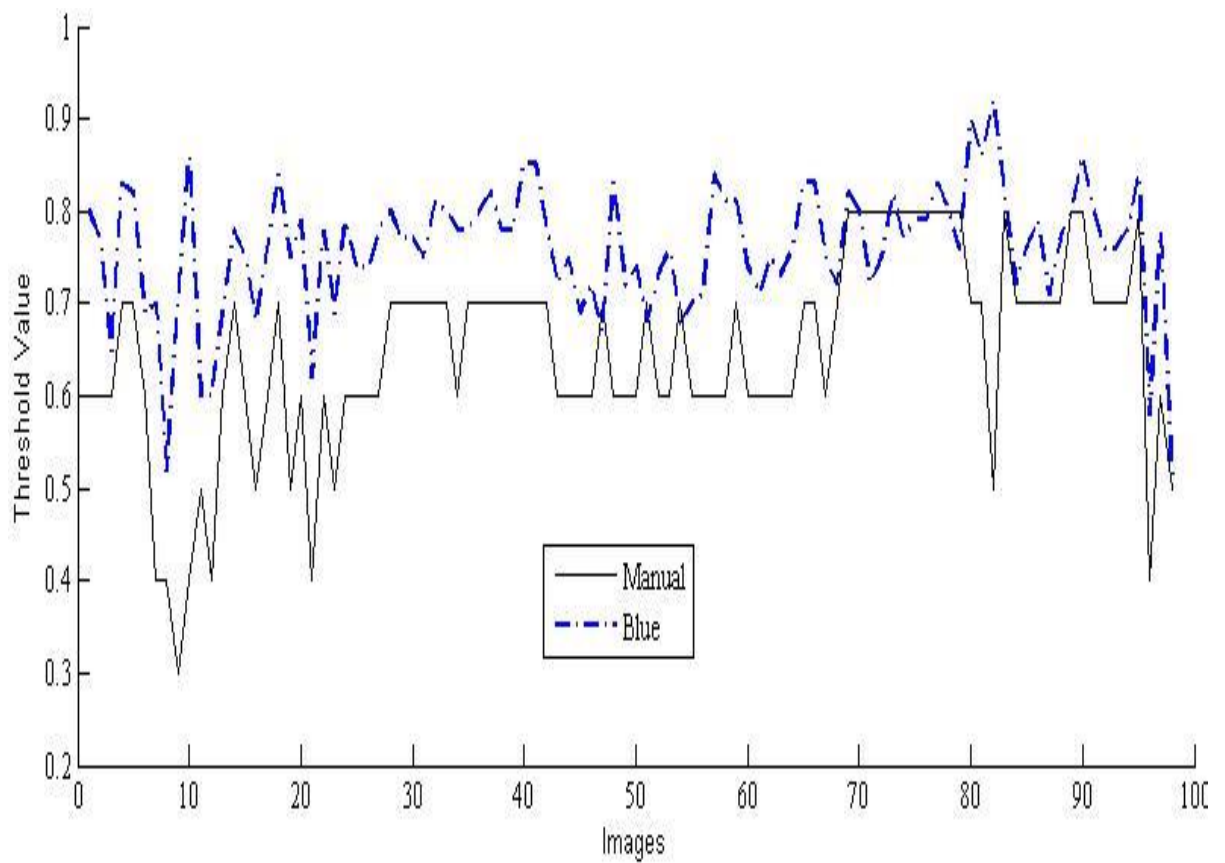


Fig 4.15: Graph of manual and blue threshold value

Images A1.7 and A1.8 from the class of Erythropoiesis both have values of green component less than the manual threshold value.

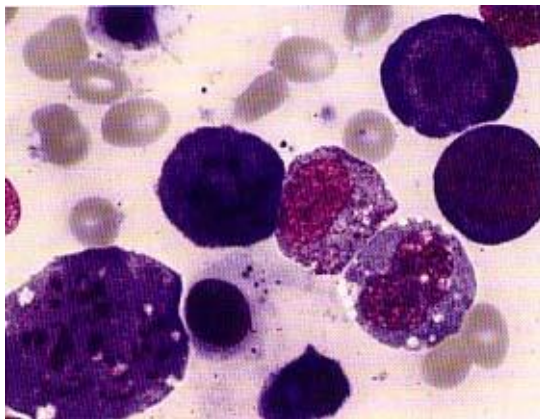


Figure 4.16: Image A1.7 from the class of Erythropoiesis

The figure 4.16 shows the image A1.7 from the class of Erythropoiesis. The cells are bigger in size and the colour is of very dark purple with two cells containing red in the center.

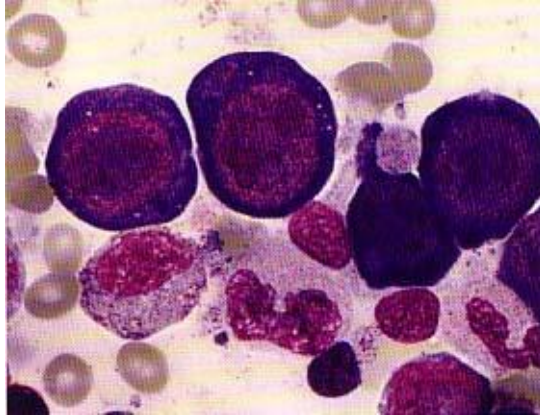


Fig 4.17: Image A1.8 from image class Erythropoiesis

For the image A1.8 (Fig. 4.17) too, the value for the green component is lower and value for the blue component is higher than the manual value. The colour of the cells for this image is very similar to image A1.7 with dark purple in the borders and red in the center.

4.4 Discussions

Four images did not produce any automatic threshold value. The result was either white or all black background. When compared with other images, these 4 images were different by content.

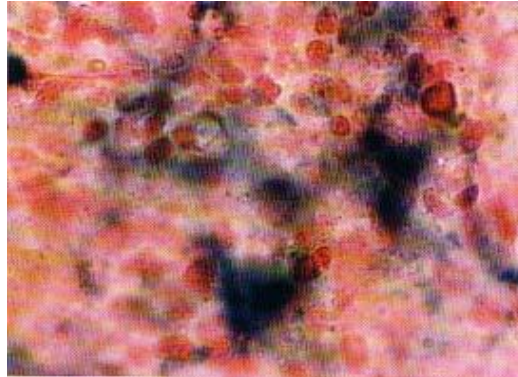


Fig 4.18: A2.1- Image from the class of Deficiency Anaemis

The content of image A2.1 (Fig. 4.18) is very muddled with dark patches here and there. There are no clear boundaries of the cells. The same applies for image A2.6 (Fig. 4.19) whereby no clear boundaries of the cells can be seen even though the image is somewhat clearer than A2.1.

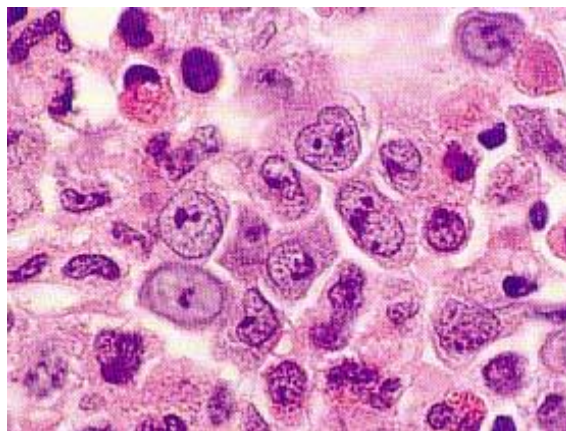


Fig 4.19: A2.6- Image from the class of Deficiency Anaemis

The other two images that could not be segmented are from Myeloid cells. Images B1.29 and B1.31 both gives a white background when segmented.

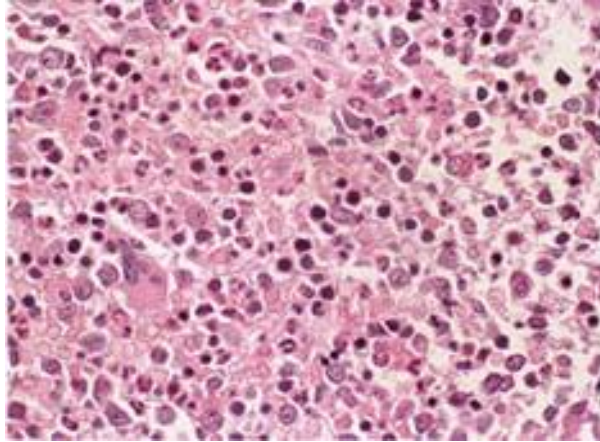


Fig 4.20: B1.29- Image from the class of Myeloid cells

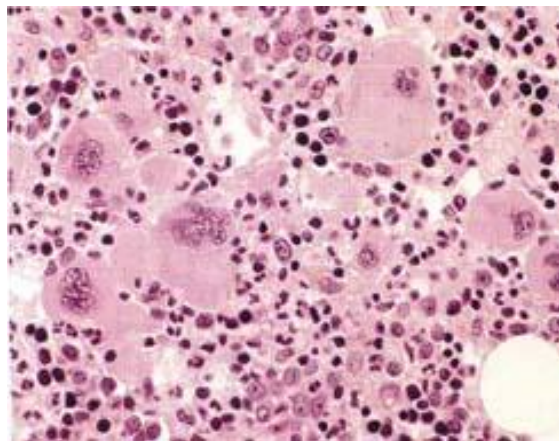


Fig 4.21: B1.31- Image from the class of Myeloid cells

Both the images appear to be very similar. The cells are not clear with small dots.

Table 4.8: Percentage of images that has a positive result

Image Classes	Number of images	Number of images with positive result	Percentage
Erythropoiesis	12	8	67
Deficiency Anaemis	6	4	67
Hemolytic Anaemis	21	20	95
Haemoglobin Disorder	18	17	94
Red cell Membrane Disorder	12	11	92
Myeloid cells	31	22	71

All 6 classes of blood cell images yield a good result. The blood cell images from the class of Hemolytic Anameis shows the highest positive result with 95%. Both the image classes of Erythropoiesis and Deficiency Anaemis have lower results among all the other classes that are 67%.

Out of 100 images, 91 images have automated and manual threshold values that is gives a difference of less than 10% that is 91%.

4.5 Summary

One hundred images from a collection of normal and abnormal blood cells have been selected as the data set. The blood cells are specifically red blood cells and are can be classified as Erythropoiesis, Deficiency anaemis, Haemolytic anaemis, Haemoglobin disorder, Red cell Membrane Disorder, Myeloid cells.

A table with differences of manually chosen threshold value and automated selection of threshold value is produced. Based on the table, a line graph is plotted to show the differences clearer. From the graph, it can be seen that the automated threshold value is higher than the manually chosen threshold value.

Four images did not give any automatic threshold value. The cells in the images were not clear and muddled. Among the 100 images, 91 images had less than 10% difference between the manual and automated threshold value, which is 91%.

Therefore, our segmentation method has produced good results while maintaining a low complexity in computation.

CHAPTER V

CONCLUSIONS

In this chapter, an overview of the whole study is looked into. The objectives of the project with the problems encountered are mentioned. Some future enhancements to the study are also suggested.

5.1 Objectives of Project

The main objective of the project was to create a system that is able to classify unknown digital images of blood cells automatically. Unclassified blood cell image is entered into the system as a query image. This image is then compared with images stored in the database. The image, which is similar to the query image in its content, is then retrieved. So, the process of classification and retrieval of the blood cell image is automated. This is beneficial to the hematologist because it saves time, human efforts and time.

The second objective is for the system to cater for different types of diseases. Blood cell images were selected from the Class Erythropoiesis, Hemolytic Anaemias, Deficiency Anaemias, Hemoglobin Disorder, Red cell Membrane Disorder and Myeloid cells. There was a high recall rate from all 6 classes of blood cells. Therefore, the system is able to cater for different types of blood cells.

The third objective was to create a simplified algorithm where intensive computation is not required. In this study, the algorithm is quite simple compared with other studies and does not require intensive computation. Although our algorithm considers only the colour

histogram of images, the experiment results were very good with a correct retrieval rate of 91%.

Another objective of this thesis was to be able to identify unknown blood cell images accurately that is having a high precision. Any query image can be classified into the 6 different classes.

The final objective of this work was to retrieve images from the database, which is highly similar to the queried image for high recall. Among the 100 images, there are 91 images with manual and automated threshold value with the difference of less than 10%. So, the recall rate is not 100%.

In summary, we had achieved the project objectives specified in Chapter 1.

5.2 Strengths of Our Method

The study does not concentrate on a particular blood cell disorder such as leukaemia cells or malarial parasite. Instead it applies to a more general blood cell images whether it is normal or abnormal. Therefore the method can cater for various types of diseases.

The method used is not complicated and does not take up computational time or cost. For example, the time taken for processing of one image by 280 x 220 pixels takes 5.83 seconds. This is considerably fast given the machine specification mentioned in section 4.1. The speed of this algorithm should be faster when implemented on a powerful machine such as Intel Core 2 Duo with 2GB of RAM.

5.3 Future Enhancements

Among the 100 images used, four images did not give any result after automated segmentation. This is because the images were more muddled with no clear cells. Detail discussions were presented in Chapter 4. For this type of images, there should be a method for users to manually select the threshold value for segmentation.

In this thesis, only one feature was extracted and that is colour. This feature alone is not adequate to differentiate each blood cell perfectly. Shape is a very important feature in distinguishing the classes of blood cells. For example, sickle cell from the class of deficiency anaemias is determined from the sickle like shape of the blood cells.

Another feature of much use is texture. A more precise descriptor is required to distinguish each of the images, as the homogeneity of the images is high. By incorporating all these features, the blood cells can be classified more accurately. However, additional features would require more complex algorithms and therefore more processing time. There is a trade-off between algorithm simplicity and retrieval accuracy.

5.4 Problems Encountered

Several problems were encountered during the study. The homogeneity of the images was one major problem. Blood cell images were homogenous in shape and colour and not clear to the naked eye. Difficulty arises in differentiating one image from the other. This requires complex method to retrieve the required images from a database.

The expertise of a haematologist is required to determine the classification of the images. Even though the images were labelled in classes, it was difficult to differentiate between images within a class itself.

There were no clear boundaries of the blood cells seen during the image segmentation process. Some of the images had overlapping cells.

5.5 Summary

This study concentrates only on images of blood cell especially on normal and abnormal red blood cells. Whether this study is applicable in other domains depends on the images used. Importantly the images have to be colour images with R, G, B components. For example, in the field of radiology, it would not be applicable as the images are grayscale images.

A limitation of the study is that other features such as shape and texture could not be incorporated, due to time constraint. So the project scope had to be narrowed down to only one feature. Only 100 images were used due to the same factor and the blood cell disorders selected were also general with no concentration on a particular type of disorder. All these arise from the scope of this thesis. This thesis is part of the major of Computer Science program, which carries only 12 credit hours out of 36 credit hours. A full research program can produce a more prospective outcome.

Content-based image retrieval is a field, which is developing rapidly, and benefiting several areas especially the medical domain. The growth in the number of medical images

produced in the medical departments has increased the necessity of having a better way of managing and retrieving these images.

In the field of haematology, a better way of classifying and retrieving images of blood cell is required. This is to increase the accuracy and effectiveness of the classification and retrieval process. A number of researches have been conducted on classifying abnormal blood cells such as leukaemia cells and malaria parasites. However not many research work in this area has been reported in the literature.

For this study, the method used was to segment the blood cells once the image was pre-processed. The pre-processing involves removal of noise and separating the image into blue and green components. Histogram, for the colour components were plotted and a threshold value was selected.

Segmentation was carried out based on the threshold value. This segmented image was then subtracted from the blood cell images, which are manually segmented and stored in the database. Images, which returns the least difference were retrieved and displayed. Among the 100 images tested, only four images could not be segmented due to the extreme content deviation of the images. Overall, the objectives of the study have been achieved.

Bibliography

- [1] F. H. Long, H. J. Zhang and D. D. Feng, "Fundamentals of content-based image retrieval," in *Multimedia Information Retrieval and Management - Technological Fundamentals and Applications*, 2003, pp. 1-26.
- [2] J. Eakins and M. Graham, "Content-based Image Retrieval", JISC Technology Applications Programme Report 39, 1999.
- [3] H. Müller , N. Michoux , D. Bandon and A. Geissbuhler, "A Review of Content-Based Image Retrieval Systems in Medical Applications – Clinical Benefits and Future Direction," *International Journal of Medical Informatics*, vol. 73, pp. 1-23, February 2004.
- [4] A. P. Sarvazyan, F. L. Lizzi, and P. N. T. Wells, "A new philosophy of medical imaging," in *Medical Hypotheses*. vol. 36, 1991, pp. 327-335.
- [5] T. Pun, G. Gerig, and O. Ratib, "Image analysis and computer vision in medicine," *Computerized Medical Imaging and Graphics*, vol. 18, pp. 85-96, 1994.
- [6] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Analysis of infected blood cell images using morphological operators," *Image and Vision Computing*, vol. 20, pp. 133-146.

- [7] J. Angulo and J. Serra, "Morphological Color Size Distributions For Image Classification and Retrieval," in *Proceedings of ACIVS 2002 (Advanced Concepts for Intelligent Vision Systems)*, Ghent, Belgium, September 9-11,2002.

- [8] Y. Rui and T. S. Huang, "Image Retrieval: Current Techniques, Promising Directions and Open Issue," *Journal of Visual Communication and Image Representation*, vol. 10, pp. 39-62, 1999.

- [9] G. Aggarwal, A. T. V, and S. Ghosal, "An Image Retrieval System With Automatic Query Modification," *IEEE Transactions on Multimedia*, vol. 4, pp. 201-214, June 2002.

- [10] QBIC(TM) -- IBM's Query By Image Content
<http://www.qbic.almaden.ibm.com/>
 Last Accessed: 7 May 2008

- [11] C. W. Niblack et al., "The QBIC project: Querying images by content using color, texture and shape," in *Proc. SPIE Storage and Retrieval of Image and Video Databases*, 1993,vol 1908, pp. 173-181.

- [12] R. B. J. Ashley, M. Flickner, J. Hafner, D. Lee, W. Niblack, and D. Petkovic, "Automatic and semi-automatic methods for image annotation and retrieval in QBIC," in *Proc. SPIE Storage and Retrieval Image and Video Databases III*, San Jose, CA, Feb. 9/10 1995, pp. 24–35.

- [13] R. P. A. Pentland, and S. Sclaroff, "Photobook: Tools for content-base manipulation of image databases," in *Proceedings of SPIE Conference on Storage and Retrieval of Image and Video Databases II*, San Jose, California, February 1994.
- [14] C. F. Jeffrey et al., "Virage image search engine: an open framework for image management," in *Proceedings of SPIE*, March 1996, pp. 76-87.
- [15] J. R. Smith and S.-F. Chang, "Querying by color regions using the VisualSEEk content-based visual query system," in *Intelligent Multimedia Information Retrieval*, 1997.
- [16] W. Y. M. and B. S. Manjunath, "Netra: A toolbox for navigating large image databases," *Multimedia Syst.*, vol. 7, pp. 184-198, 1999.
- [17] J. L. James, Z. Wang, and G. Wiederhold, "SIMPLicity: Semantics-sensitive Integrated Matching for Picture Libraries," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 947-963, 2001.
- [18] L. S. Anca Ion, "MIR - A Tool for Medical Image Retrieval" in *The International Conference on Computer as a Tool, EUROCON 2005*, Belgrade, Serbia and Montenegro, 2005, pp. 1003 - 1006.
- [19] J. R. Smith and S.-F. Chang, "Automated Image Retrieval Using Color and Texture," Columbia University, Technical Report CU/CTR 408-95-, 14 July 1995.

- [20] D. Zhang, "Improving Image Retrieval Performance by Using Both Color and Texture Features," in *Proceedings of the Third International Conference on Image and Graphics (ICIG'04)*.
- [21] X. Y. C. Pan, and C. Zheng, "Recognition of Blood and Bone Marrow Cells using Kernel-based Image Retrieval," *International Journal of Computer Science and Network Security*, vol. 6, Oct 2006.
- [22] Y.-K. Chan and C.-Y. Chen, "Image retrieval system based on color-complexity and color-spatial features," *Journal of Systems and Software*, vol. 71, pp. 65-70, April 2004.
- [23] S. Jeong, "Histogram-Based Color Image Retrieval," Psych221/EE362 Project Report, 15 Mar 2001.
- [24] S. Wang, "A Robust CBIR Approach Using Local Color Histograms", October 2001,
<http://citeseer.nj.nec.com/wang01robust.html>
Last Accessed: 10 February 2008
- [25] Md. Mahmudur Rahman, T. Wang, and B. C. Desai, "Medical Image Retrieval and Registration: Towards Computer Assisted Diagnostic Approach" in *Medical Information Systems: The Digital Hospital, 2004. IDEAS '04-DH. Proceedings. IDEAS Workshop on*, 1-3 Sept. 2004, pp. 78 - 89.

- [26] H. Greenspan and A.T. Pinhas, "Medical Image Categorization and Retrieval for PACS Using the GMM-KL Framework," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, pp. 190 - 202, March 2007.
- [27] S. Dube, S. El-Saden, T. F. Cloughesy, and U. Sinha "Content Based Image Retrieval for MR Image Studies of Brain Tumors," in *Engineering in Medicine and Biology Society. 28th Annual International Conference of the IEEE* 2006, pp. 3337 - 3340.
- [28] S. K. Kinoshita, P. M. D. Azevedo-Marques, R. R. Pereira, J. A. H. Rodrigues, and R. M. Rangayyan, "Content-based retrieval of mammograms using visual features related to breast density patterns", *Journal of digital imaging*, vol. 20, pp. 172-90, Jun 2007.
- [29] Y. Tao, S.-C Lo, MT Freedman, and J. Xuan, "A preliminary study of content-based mammographic masses retrieval," in *Proc. SPIE*, 2007.
- [30] A. W. L Zheng, J Gilbertson, and M. J. Becich "Design and analysis of a content-based pathology image retrieval system" *IEEE Trans Inf Technol Biomed*, vol. 7, pp. 249-55, 2003.
- [31] S. Antania, D.J. Leeb, L. R. Longa, and G. R. Thomaa, "Evaluation of shape similarity measurement methods for spine X-ray images" *Journal of Visual Communication and Image Representation*, vol. 15, pp. 285-302, September 2004.

- [32] Image Retrieval in Medical Application (IRMA)
http://phobos.imib.rwth-aachen.de/irma/index_en.php
Last Accessed: 20 April 2008
- [33] J. D. D. Keyzers, H. Ney, B.B. Wein, and T.M. Lehmann, "A statistical framework for model-based image retrieval in medical applications," *J. Electronic Imag.* , vol. 12, pp. 59-68, 2003.
- [34] B. W. Thomas, M. Lehmann, J. Dahmen, J. Bredno, F. Vogelsang, and Michael Kohlen, "Content-Based Image Retrieval in Medical Applications: A Novel Multi-Step Approach."
- [35] H. L. Tang and R. Hanka, and H. S. Horace, "Histological Image Retrieval Based on Semantic Content Analysis" *IEEE Transactions o Information Technology in Biomedicine*, vol. 7, March 2003.
- [36] I-Browse - An Image Retrieval and Annotation System Based on Semantic Content
<http://icg.cityu.edu.hk/ICGers/KentCheung/ibrowse/ibrowse.htm>
Last Accessed: 20 Feb 2008
- [37] W. W. Chu, A. F. C. Ardenas, and R. K. Taira, "KMED: A knowledge-based multimedia distributed database system," *Information Systems*, vol. 19, pp. 33-54, 1994.

- [38] The Knowledge-Based Multimedia Medical Distributed Database System
<http://kmed-www.cs.ucla.edu/intro.html>
Last Accessed: 15 April 2008
- [39] M. E. Mattie, L. Staib, E. Stratmann, H. D. Tagare, J. Duncan, and P.L. Miller, "PathMaster: Content-based Cell Image Retrieval Using Automated Feature Extraction," *J Am Med Inform Assoc.*, vol. 7, pp. 404-415, Jul/Aug 2000.
- [40] C.R. Shyu, C. Brodley, A. Kak, A. Kosaka, A. Aisen, and L.S. Broderick, "ASSERT: A Physician-in-the-loop Content-Based Retrieval System for HRCT Image Databases," *Computer Vision and Image Understanding*, vol. 75, pp. 111-132, 1999.
- [41] A. M. Aisen et al., "Automated Storage and Retrieval of Thin-Section CT Images to Assist Diagnosis: System Description and Preliminary Assessment," *Radiology*, vol. 228, pp. 265-270, July 2003.
- [42] H. Muller, A. Rosset, J. P. Vallee, and A. Geissbuhler, "Comparing feature sets for content-based image retrieval in a medical case database" in *Proceedings of SPIE*, 2004, vol 5371, pp 99-109.
- [43] H. Muller, J. Heuberger, A. Depeursinge, and A. Geissbuhler, "Automated Object Extraction for Medical Image Retrieval Using the Insight Toolkit (ITK)", in *Proceedings of Information retrieval technology*, October 16-18, 2006

- [44] H. Müller, A. Rosset, A. Garcia, J.-P. Vallée, and A. Geissbuhler, "Benefits of Content-based Visual Data Access in Radiology," *RadioGraphics*, vol. 25, pp. 849-858, 2005.
- [45] E.A. El-Kwae, H. Xu, and M.R. Kabuka, "Content-based retrieval in picture archiving and communication systems," *J Digit Imaging*, vol. 13, pp. 70-81, May 2000
- [46] S.C. Orphanoudakis, C. Chronaki, and S. Kostomanolakis, "I2C: a system for the indexing, storage, and retrieval of medical images by content," *Med Inform (Lond)*, vol. 19, pp. 109-22, Apr-Jun1994
- [47] Guangquan Zhou, Liang Ju, Limin Luo, Xudong Bao, and Huanzhong Shu, Medical Imaging 2004:PACS and Imaging Informatics, "Content-based cell pathology image retrieval by combining different features," in *Proceedings of SPIE.*, April 2004, vol. 5371, pp. 326-333.
- [48] M. S. Nixon, and A. S. Aguado, "Feature Extraction and Image Processing", First edition 2002 ed.
- [49] D. Sagarmay and Z. Yanchun, "An Overview of Content-based Image Retrieval Techniques," in *Proceedings of the 18th International Conference on Advanced Information Networking and Application (AINA '04)*.
- [50] P. R.Wheater, H. G. Burkitt, and V.G. Daniels, *Fuctional Histology*: ELBS, 1987.

- [51] P. Kumar and M. Clark, *Clinical Medicine*: Elsevier Saunders, 2005.
- [52] Blood Cells
http://www.funsci.com/fun3_en/blood/blood.htm
Last Accessed: 20 March 2008
- [53] MedlinePlus Medical Encyclopedia
<http://www.nlm.nih.gov/medlineplus/ency/article/000584.htm>
Last Accessed: 30 March 2008
- [54] emedicine
<http://www.emedicine.com/med/topic979.htm>
Last Accessed: 30 March 2008
- [55] World Health Organization
<http://www.who.int/mediacentre/factsheets/fs308/en/index.html>
Last Accessed: 25 March 2008
- [56] S. E. Umbaugh, *Computer Vision and Image Processing: A practical approach using CVIP tools*: Prentice Hall, 1998.
- [57] Qiang Wang, V. Megalooikonomou, and D. Kontos, "Medical Image Retrieval Framework," in *2005 IEEE Workshop on Machine Learning for Signal Processing*, pp. 233 – 238.

- [58] A.W.M. Smeulders, M.Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years ", vol. 22, pp. 1349-1380, December 2000.
- [59] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age ", *ACM Computing Surveys*, vol. 40, p. 60, 2008.
- [60] T.-S. Chua, S.-K. Lim, and H.-K. Pung, "Content-based retrieval of segmented images," in *International Multimedia Conference Proceedings of the second ACM international conference on Multimedia*, San Francisco, California, United States, pp. 211 - 218.
- [61] D. Anoraganingrum, "Cell Segmentation with Median Filter and Mathematical Morphology Operation," in *10th International Conference on Image Analysis and Processing (ICIAP'99)*, 1999, p. 1043.
- [62] Y. A. Asl. and C. T. Yu, "Techniques and Systems for Image and Video Retrieval", *IEEE Transactions on Knowledge and Data Engineering* vol. 11, pp. 56-63, January/February 1999.
- [63] P. Howarth, A. Yavlinsky, D. Heesch, and S. Ruger, "Medical Image Retrieval Using Texture, Locality And Colour," in *CLEF 2004, 5th Workshop of the Cross-Language Evaluation Forum*, C. P. Peters C, Gonzalo J Ed. Bath, UK, 2004, pp. 740-749.

- [64] P. Howarth, A.Yavlinsky, D. Heesch, and S. Ruger, "Medical image retrieval using texture, locality and colour," *Lect Notes Comput Sci*, vol. 3491, pp. 740-749, 2005.

- [65] T. M. Lehmann et.al., "Content-based image retrieval in medical applications for picture archiving and communication systems," in *Proceedings of IS&T/SPIE Medical Imaging 2003: PACS and Integrated Medical Systems*, February 2003.

- [66] M.O. Güld, C. Thies, B. Fischer, and T. M. Lehmann, "A generic concept for the implementation of medical image retrieval systems," *International Journal of Medical Informatics*, vol. 76, pp. 252-259, February-March 2007.

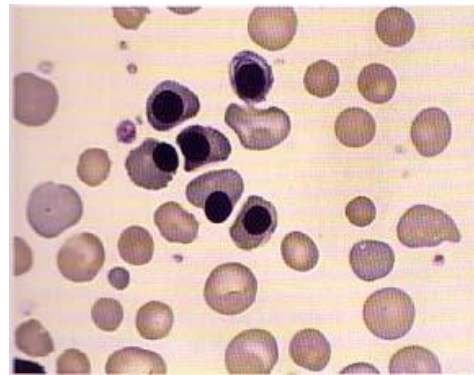
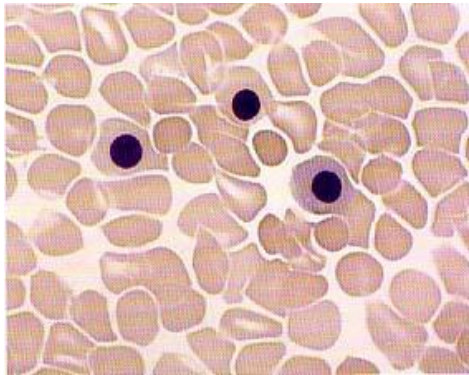
- [67] W. Cirne, M.C. Oliveira, and Paulo M. de Azevedo Marques, "Towards applying content-based image retrieval in the clinical routine ", *Future Generation Computer Systems*, vol. 23, pp. 466-474, March 2007.

- [68] I. El-Naqa, Y.Yang, N.P. Galatsanos, R.M. Nishikawa, and M.N. Wernick, "A similarity learning approach to content-based image retrieval: application to digital mammography.," *IEEE Trans Med Imaging*, vol. 23, pp. 1233-44, Oct 2004

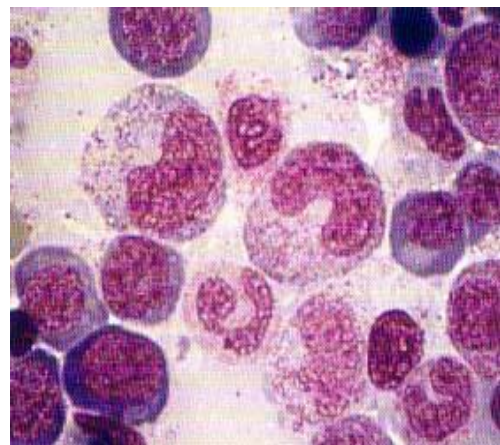
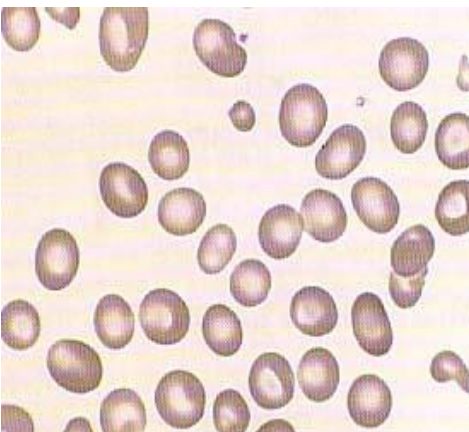
Appendix

Blood cell sample

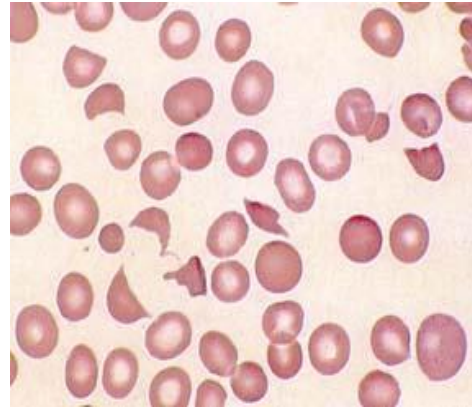
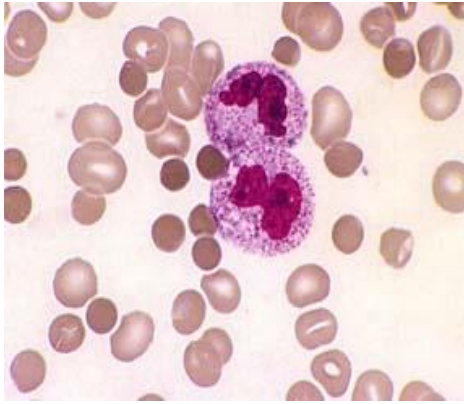
Erythropoiesis



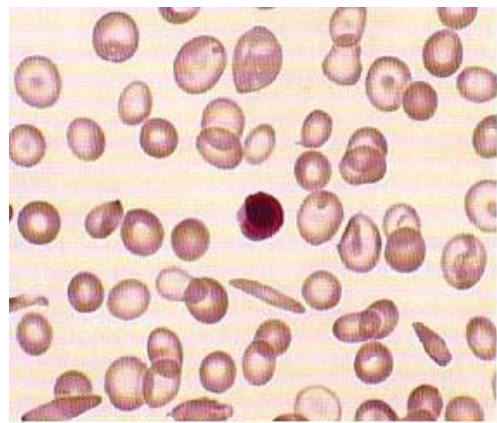
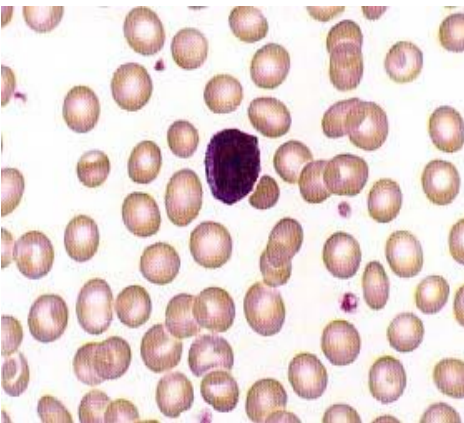
Deficiency Anaemias



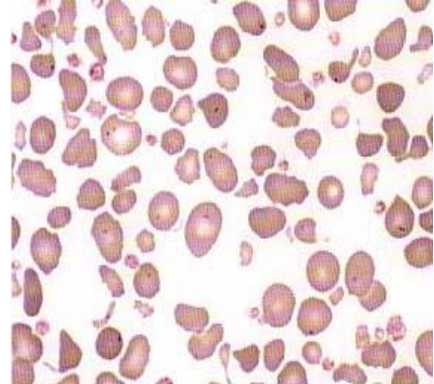
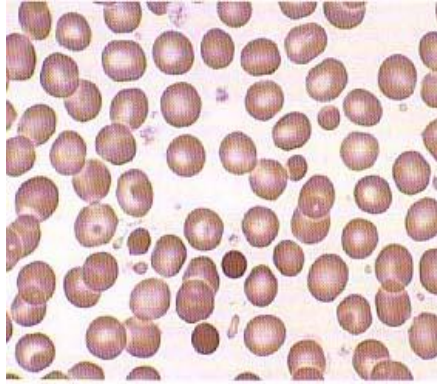
Haemolytic Disorder



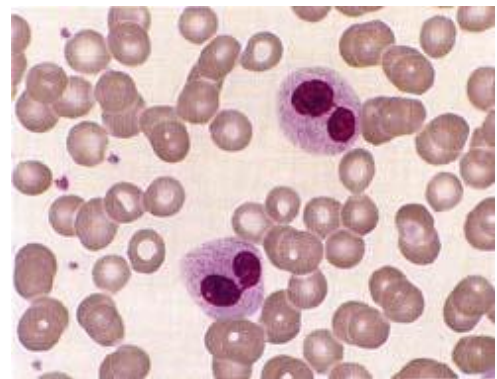
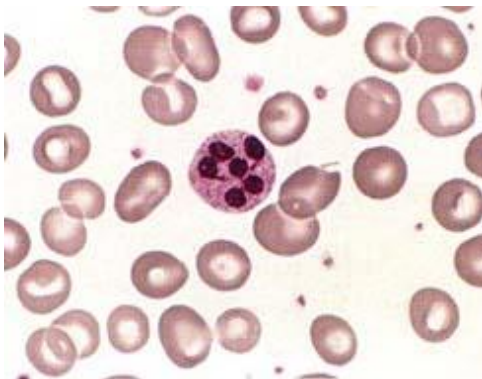
Haemoglobin Disorder



Red Membrane Cell Disorder



Myeloid Cells



Introduction: MATLAB

MATLAB is a high-level technical computing language and has an interactive environment for algorithm development, data visualization, data analysis and numeric computation. MATLAB can be used in signal and image processing, communications, control design, test and measurement, financial modeling and analysis and computational biology. MATLAB has add-on toolboxes that extend its environment to solve particular classes of problems. MATLAB's capabilities have been enhanced with an Image processing toolbox for image processing.

MATLAB is a numerical computing environment and programming language. Created by *The MathWorks*, MATLAB allows easy matrix manipulation, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs in other languages. Although it specializes in numerical computing, an optional toolbox interfaces with the Maple symbolic engine, allowing it to be part of a full computer algebra system.

MATLAB was first adopted by control design engineers but quickly spread to many other domains. It is now also used in education, in particular the teaching of linear algebra and numerical analysis, and is popular amongst scientists involved with image processing.

Introduction: Image Processing Toolbox 5.4

The Image Processing Toolbox supports engineers and scientists in areas such as biometrics, remote sensing, surveillance, gene expression, microscopy, semiconductor testing, image sensor design, colour science, and materials science. It also facilitates the learning and teaching of image processing techniques.

The Image Processing Toolbox provides a comprehensive suite of reference-standard algorithms and graphical tools for image analysis tasks such as statistical analysis, feature extraction, and property measurement.

Statistical functions let you analyze the general characteristics of an image by:

- Computing the mean or standard deviation
- Determining the intensity values along a line segment
- Displaying an image histogram or plotting a profile of intensity values

Edge-detection algorithms let you identify object boundaries in an image. These algorithms include the Sobel, Prewitt, Roberts, Canny, and Laplacian of Gaussian methods. The powerful Canny method can detect true weak edges without being “fooled” by noise.

Image segmentation algorithms determine region boundaries in an image. You can explore many different approaches to image segmentation, including automatic thresholding, edge-based methods, and morphology-based methods such as the watershed transform, often used to segment touching objects.

Morphological operators enable you to detect edges, enhance contrast, remove noise,

segment an image into regions, thin regions, or perform skeletonization on regions.

Morphological functions in the Image Processing Toolbox include:

- Hole filling
- Peak and valley detection
- Watershed segmentation
- Reconstruction
- Distance transform

The Image Processing Toolbox also contains an advanced image analysis function that enables:

- Measuring the properties of a specified image region, such as the area, center of mass, and bounding box.
- Detecting lines and extract lines segments from an image using the Hough transform.
- Measuring properties, such as surface roughness or colour variation, using texture analysis functions.

Developing Algorithms

MATLAB and the Image Processing Toolbox provide a comprehensive platform for interacting with image data and developing image processing algorithms. The toolbox adds reference-standard image processing and analysis algorithms and image-specific graphical tools to the high-level language, basic mathematical functions, development tools, and visualization capabilities of MATLAB.